Journal of
Clinical Psychology

OCTOBER 1959
Volume XV Number 4

EDITORIAL BOARD

FREDERICK C. THORNE, *Editor*
Editorial Office: 5 Pearl Street
Brandon, Vermont

JERRY W. CARTER, JR.
U. S. Public Health Service

HOWARD F. HUNT
University of Chicago

WILLIAM A. HUNT
Northwestern University

ELAINE F. KINDER
New York State Department of Mental Hygiene

MAURICE LORR
Veterans Administration Washington, D. C.

SAUL B. SELLS
Texas Christian University

ROBERT I. WATSON
Northwestern University

ELEANOR C. THORNE, *Business Manager*

Journal of Clinical Psychology
5 Pearl Street · Brandon, Vermont

PSYCHOLOGICAL TEST SPECIALISTS

Exclusive publishers and distributors for selected psychological tests

TESTS NOW AVAILABLE

Kahn Test of Symbol Arrangement

Highly versatile projective test using symbolic plastic objects which are arranged and sorted. Validly discriminates among normals, neurotics, brain-injured, and psychotics, and indicates accompanying personality dynamics. **$25.00 for set of materials and manuals**

Full-Range Picture Vocabulary Test

Brief test of intelligence in two forms using the same 16 plates. Words are defined by choosing pictures. Valuable for intelligence screening, or with the physically handicapped. **$7.50 for set of materials**

Proverbs Test

Group and individual forms determine similarity of thought processes to those of various normal and clinical groups. Highly sensitive indicator of psychotic processes. **$4.50 for kit**

Vocational Apperception Test -- College Form

Separate male and female series of plates depicting college-related occupational activities, accompanied by brief explanation of use and interpretation as a projective test. **$7.25 for set of materials**

Famous Sayings

Agreement with proverbs, aphorisms and folk sayings is analyzed to determine personality structure. Of particular use in personnel selection. **$4.50 for kit**

Geist Picture Interest Inventory

Provides for expression of interests by choices among alternative pictured activities. Both qualitative and quantitative interpretations are possible. **$1.00 for sample set of materials**

Quick Picture Vocabulary Test

Each of three forms based on one test plate gives a good estimate of intellectual ability in 2-3 minutes. Can be used independently or as third form of FRPV. **$5.00 for set of materials**

Id-Ego-Superego Test

Subtests give independent estimates of personality components from a modified psychoanalytic point of view. **$28.50 for set of materials and manual**

TESTS TO BE PUBLISHEDIN THE IMMEDIATE FUTURE

Emotional Projection Test

Emotional states are attributed to persons shown in 30 photographs. Extensive research has shown that responses to this test are extremely sensitive indicators of needs and feelings and can be effectively analyzed to determine personality structure.

Group Personality Projective Test

A multiple-choice projective test using 90 cartoon-like drawings in a booklet allows quick objective analysis of a variety of personality characteristics.

TESTS IN PREPARATION

Vocational Apperception Test-General Form; Electron Micrograph Apperception Test; It Scale for Children; Projective Movement Sequences; Polychrome Index; Memory for Designs Test.

To order any of the tests now available, or to obtain information about those to be published, write:

PSYCHOLOGICAL TEST SPECIALISTS

Box 1441 MISSOULA, MONTANA

JOURNAL OF CLINICAL PSYCHOLOGY

VOL. XV | OCTOBER 1959 | No. 4

CONTENTS

STUDIES OF PERSONALITY STRUCTURE

Personality Structure in Four and Five Year Olds in Terms of Objective Tests. RAYMOND B. CATTELL and DONALD R. PETERSON - - - - 355
The Meaning of a Positive Self-Concept. FREDERICK M. JERVIS - - - 370
Difference between Neurotics, Psychotics and Normals on Perceptual Tests: Reanalysis of Eysenck's Results. STEVEN G. VANDENBERG - - - - 373
Homosexuality and Creativity. ALBERT ELLIS - - - - - - - - 376

STUDIES OF CLINICAL JUDGMENT PROCESSES

Judging the Relative Severity of Psychiatric Out-Patient Complaints. K. WARNER SCHAIE, LOIS A. ROMMEL and JAMES M. A. WEISS - - - 380
The Comparative Reliability of Two Techniques of Personality Appraisal. PAUL HORST and CALVIN E. WRIGHT - - - - - - - - - 388
Reliability of Diagnostic Judgments Made by Psychologists. ALLAN GOLDFARB - 392
The Validity of Clinical Judgments of Schizophrenic Pathology Based on Verbal Responses to Intelligence Test Items. NELSON F. JONES, JR. - - 396
A Comparison of Random and Judgmental Methods of Determining Mode of Outpatient Mental Hygiene Treatment. FLOYD K. GARETZ, RICHARD L. KOGL and DANIEL N. WIENER - - - - - - - - - - - - 401
A Comparison of Parental and Self-Evaluations of Psychopathology in Children. EUGENE E. LEVITT - - - - - - - - - - - - 402

DIAGNOSTIC PROCESSES: STANDARDIZATION, VALIDATIONAL and NORMATIVE STUDIES

Standardization of the Children's Picture Information Test. KATE L. KOGAN and RICHARD L. CRAGER - - - - - - - - - - - - 405
The Pain Apperception Test: An Application to Sex Differences. DONALD V. PETROVICH - - - - - - - - - - - - - - - - - 412
An Analysis of Gross Errors on a Sentence Completion Test. WILSON H. GUERTIN - 415
Deviant Responses as Indicators of Immaturity and Neuroticism. JOHN C. ROITZSCH and IRWIN A. BERG - - - - - - - - - - - - 417
An Internal Consistency Analysis of the Forced Choice Anxiety Scale. JOHN A. HAMMES and E. DELL YOUNG - - - - - - - - - - - 420
Formboard Comparability in the Colored Progressive Matrices. THOMAS E. JORDAN - 422
A Short Form of the Wechsler Intelligence Scale for Children. WILLIAM H. SIMPSON and CECIL C. BRIDGES, JR. - - - - - - - - - - - 424
Base Rate Data on Rorschach Card Rejection. WALTER F. MCKEEVER and ALVIN I. GERSTEIN - - - - - - - - - - - - - - - - 425
MMPI Profiles of Parents of Child Psychiatric Patients. LEO J. HANVIK and MILDRED BYRUM - - - - - - - - - - - - - - - - 427

CONTENTS—Continued

THE EFFECTS OF DRUGS ON MENTAL PROCESSES
Changes in Recognition Thresholds Associated with Chlorpromazine, Promazine and Phenobarbital. HERMAN Y. EFRON - - - - - - - 431
Effects of Morphine and Pentobarbital on Differential MMPI Profiles. CHARLES A. HAERTZEN and HARRIS E. HILL - - - - - - - - 434
Intellectual Changes Following Frontal Lobe Procainization. ALBERT F. PAOLINO and IRA FRIEDMAN - - - - - - - - - - - - - 437

Anxiety and Socio-Economic Stratification. ERWIN J. LOTSOF and RICHARD CENTERS - - - - - - - - - - - - - - - - - - 439
A Study of Factors Related to Length of Stay in Psychotherapy. M. A. BAILEY, L. WARSHAW and R. M. EICHLER - - - - - - - - 443
Multidisciplinary Views on the Preparation of Written Clinical Psychological Reports: III. The Trouble with Psychological Reports. NORMAN TALLENT and WILLIAM J. REISS - - - - - - - - - - - - - - - 444
Some Cautions Concerning the Use of Change Scores. FRED HEILIZER - - 447
The Effect of Tied Ranks on the Rank Order Correlation. FRED HEILIZER - 450
Hypochondriasis and Somatic Seizure Auras. BERNARD S. AARONSON - - 450
The Shipley-Hartford Scale and the Doppelt Short Form as Estimates of WAIS IQ in a State Hospital Population. LLOYD K. SINES and HELEN SIMMONS. 452

Editorial Opinion - - - - - - - - - - - - - - - - - - 454
Book Notices - - - - - - - - - - - - - - - - - - - 456
Table of Contents, Vol. XV, 1959 - - - - - - - - - - - - - - 457
Author Index, Vol. XV, 1959 - - - - - - - - - - - - - - - 462

GENERAL INFORMATION

The *Journal of Clinical Psychology* is an independent journal dedicated to the advancement of the clinical method in psychology. Although primarily a scientifically oriented professional journal limited to publication of original research reports and authoritative theoretical articles, it aims to foster the promotion and expansion of clinical psychology as an applied science. Communications and articles submitted for publication are welcome and should be addressed to the Editorial Office, 5 Pearl Street, Brandon, Vermont.

ORIGINAL ARTICLES are published only with the understanding that they are contributed exclusively to the *Journal of Clinical Psychology*. The publishers are not responsible for statements made or opinions expressed by contributors in articles printed in these columns. Copyrights cover publication in the *Journal of Clinical Psychology* and articles may not be reproduced without permission of the publishers.

THE STYLE of all manuscripts including bibliographies must be in accordance with the instructions given in: Anderson, J. E., and Valentine, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 41, 345-376. References should be arranged alphabetically according to the directions given above. Full address of the author should appear somewhere in the article, preferably at the end.

REPRINTS are furnished on order only and must be requested on the order blank provided when galley proofs are returned. Change of address notices should be sent one month before moving and should include old and new addresses. Undelivered copies will be remailed only at subscriber's expense.

The *Journal of Clinical Psychology* is published quarterly, appearing in January, April, July and October, at Rutland, Vermont, by Frederick C. Thorne, M.D. Subscription price $10.00; Canadian subscriptions $10.50; foreign subscriptions $11.00 including postage; single copies $2.50. Entered as second class matter January 15, 1945, at the post office at Burlington, Vermont, under the Act of March 3, 1879. Reentered at the post office at Brandon, Vermont. Additional entry at the post-office at Rutland, Vermont. *Title registered, U. S. Patent Office.* Editorial and Business Offices, 5 Pearl Street, Brandon, Vermont. Second-class postage paid at Rutland, Vt. and at additional mailing offices.

Copyright, 1959, Frederick C. Thorne, M. D.

PERSONALITY STRUCTURE IN FOUR AND FIVE YEAR OLDS IN TERMS OF OBJECTIVE TESTS

RAYMOND B. CATTELL AND DONALD R. PETERSON

University of Illinois

PROBLEM

The importance for clinical psychology of the first few years of life has been fully conceded, and endlessly discussed, but until measurement of temperamental tendencies and dynamic structural growth in this period can be achieved, little progress can be made in exactly checking and refining existing theories.

Multivariate, factor analytic experiment has been successful in confirming clinical structures at the adult [6] and middle childhood [8, 9, 10, 11, 12] levels, but a fear that personality structure might be too fluid and subtle at the four year level seems hitherto to have deterred systematic research there. However, the present writers have shown that parent [18] and nursery school teacher [17] behavior ratings are structured by factor analysis with considerable certainty and uniqueness, and yield patterns which seem to be recognizably the precursors of the ego strength dimension, the schizothyme temperament, surgency-desurgency, *etc.*, at later ages. Work in this area has been further encouraged by the findings of reliable patterns in verbal questionnaire responses [16] and in time samplings of actual behavior recorded in the playground situation [12].

However, if these structures are to become the pivots of experimental work on child personality development, it is necessary to shift their measurement from cumbersome rating procedures on to a basis of easily obtainable objective tests, and it is with the development of such tests that the present contribution is concerned. By objective tests we mean not merely score-objective, "conspective" [6] tests as in a questionnaire, but, rather, that variety of situational, misperceptive ("projective"), stylistic, etc., tests as defined elsewhere under "objective" [7]. Parenthetically, the present study is part of a larger, long-term research exploring personality structure by objective means through all three media of observation (life behavior *in situ*, questionnaire and objective test) and at all age levels, an interim integration of which has recently appeared [6].

Obviously much is going to turn, in the present endeavor, on getting tests which get out into the full space of the *personality sphere* [6]. The invention and choice of tests at this level has benefited from experience at other age levels and followed the same canons, namely: (a) To include a sufficiency of unchanged tests, *to mark* (as salients) *the factors* found in the above preceding and coordinated studies; and (b) to shape *new* tests to get at the forms of response already indicated in the behavior rating and questionnaire factorizations at this age to be important in covering comprehensively the personality sphere. The first is most important in the initial exploration at a new age level, for, unless older, known factors can be "placed", there is confusion. But, once they have been so placed, new test invention can proceed, in order to yield, in later experiments, higher validities (loadings) for the now recognized and confirmed patterns.

With such a basis of test selection, our design became: (a) to administer the battery to a sufficient sample of 4-5 year olds, (b) to discover the number and nature of the factors accounting for the correlations, by pursuing simple structure, if found inherent in the data, blindly (*i.e.*, by statistical criteria alone) and uninfluenced by theoretical preconceptions, and, (c) to develop sufficient precision of hypotheses on these factors, *e.g.*, through interpretation in terms of factors known at other ages, to make a basis for further checking researches. The same children tested here on objective tests have also been subjects for factoring on ratings (including time sampling) [13, 16] and on an oral questionnaire [15], and the resulting factors from the three media are later to be brought together.

The Nature of the Personality Tests

As just indicated, it was our intention to represent each personality factor hitherto known, and specifically to represent each factor, old or newly-hypothesized by *at least two marker tests*, as close as possible to the tests used at older levels, *i.e.*, these particular tests should be of a kind which one would not expect to be modified very much in going down the scale from eleven and seven to four years of age. By this means, after rotation, we might hope to identify the factors with the factors of the upper age level. In addition to these markers, a number of entirely new tests were invented especially adapted to the four year old level and concerned with behavior which might clearly bring out the meaning of certain factors. The resulting list of variables is given in Table 1.[1]

Table 1. List of Objective Test Variables

The following 81 variables are indexed in order of appearance in the correlation and factor matrices, as well as with a Master Index (M.I.) number permitting their location. For full particulars of administration and scoring, see the Test Compendium, the O-A Battery [5] and other articles and books [6, 9, 12, 14] describing the tests. The verbal title is kept as close to the habitual title in laboratory publications as possible (allowing for slight changes in descriptive emphasis in referring as accurately as possible to the child form), and the adjective at the right indicates the addition for sign of loading in the factor salient tables.

The division into "main variables" and "secondary variables" is of no major important, but the former were less "speculative" and the latter were often only modifications of scoring of earlier tests deemed more essential to defining factor structure.

No. of Variable in Factor Matrix	Master Index Number	Main Variables Title to be given to Positive Loadings	High Test Score	Low Test Score	Reliability
1.	M.I. 176	Regularly warned reaction time (R.T.)	Long	Short	.79
2.	M.I. 176	Irregularly warned reaction time (R.T.)	Long	Short	.93
3.	M.I. 40	Reactions to false signals (R.T.)	Many	Few	.52
4.	M.I. 5	Ratio of irregular to regular warned reaction time	Long	Short	.16
5.	M.I. 263	Variance of simple reaction time	Large	Small	.56
6.	M.I. 335	Complex decision reaction time	Long & Accurate	Short & Inaccurate	.75
7.	M.I. 71	Errors of decision in maintaining complex reaction time sets	Many	Few	.50
8.	M.I. 72	Ratio of complex to simple reation time	Large	Small	.71
9.	M.I. 307a	Speed of decision in visual comparisons	Short	Long	—
10.	M.I. 270	Tempo of arm-shoulder movement	Many	Few	.92
11.	M.I. 282	Objects seen in unstructured drawings	Many	Few	.77
12.	M.I. 336	(*Absolute number*) threatening objects in unstructured drawings	Many	Few	.62
13.	M.I. 105	Percentage threatening objects perceived in unstructured drawings	Many	Few	.46
14.	M.I. 168	Speed of social judgment	Short	Long	
15.	M.I. 27	Willingness to play practical jokes and tease	Many jokes	Few jokes	.62
16.	M.I. 21	Questionable preferences in reading	Many emotional	Few emotional	.09
17.	M.I. 15	Use of circles and slanting lines equivalent on C.M.S. Test. (Child version)	Many	Few	.88
18.	M.I. 98	Ratio incorrect (forbidden) to total number on CMS performance (Child version)	Many correct	Few correct	.81

[1]The originals of these tests may be obtained for research purposes on application to the authors.

19.	M.I. 143a	Total (correct) score on CMS (modified) (Child version)	Many correct	Few correct	.71
20.	M.I. 30	Ratio self criticism to criticism of others	Many self criticisms	Many criticisms of others	.34
21.	M.I. 29	Criticism of performances.	Many criticisms	Few criticisms	.44
22.	M.I. 146	Accuracy of perceptual Gestalt completion	High percentage correct	Low percentage correct	.92
23.	M.I. 337	Speed of perceptual Gestalt completion	Short time	Long time	
24.	M.I. 51	Index of Carefulness	Many correct	Few correct	.11
25.	M.I. 206	Analytic performance on Gottschaldt figures	Many correct	Few correct	.46
26.	M.I. 338	Critical exactness	High percentage correct	Low percentage correct	.69
27.	M.I. 17	Tendency to agree	High percentage	Low percentage	.20
28.	M.I. 219	Willingness to admit being naughty (common frailties)	Many confessed	Few confessed	.20
29.	M.I. 38	Tendency to recall consonant vs. disonant statements	High ratio consonant	High ratio disonant	.46
30.	M.I. 332	Willingness to forego prize for those in need	Gives up many rewards	Will keep rewards	.66
31.	M.I. 150	Extremity of viewpoint	Many extreme categories	Many middle categories	.51
32.	M.I. 34	Maturity of opinion	Few changes	Many inconsistent responses	.06
33.	M.I. 271	Fluency	Many responses	Few responses	.60
34.	M.I. 6:128	Ideomotor speed (color naming)	Many named	Few named	.80
35.	M.I. 2	Rigidity (reversed color naming)	Reversed score low	Regular score low	.37
36.	M.I. 2	Performance on Rigidity (mirror drawing)	Fails to go far		.18
37.	M.I. 339	Performance on mirror drawing	Long time per unit score	Short time per unit score	.39
38.	M.I. 227	Preference for color/form in pictures	Many color	Many form	.32
39.	M.I. 120	Ratio of inaccuracy to speed (pencil maze)			.76
40.	M.I. 334	Ratio inaccuracy (to speed) on difficult to easy mazes			.09
41.	M.I. 144	Acceptance of moral aphorisms (platitudes)	Agrees with many	Agrees with few	.07
42.	M.I. 333	Readiness to learn new names	Many correct	Few correct	.49
43.	M.I. 211	Annoyability (all situations)	Many annoyances	Few annoyances (& moderate)	.64
44.	M.I. 330	Ethical choices	Many	Few	.26
45.	M.I. 47	Expansiveness of copied design	Large drawing	Small drawing	.71
46.	M.I. 342	Speed of drawing	Short time	Long time	.69
47.	M.I. 340	Speed on repetitive formboard performance (realistic manipulation)	Short time	Long time	.96
48.	M.I. 199	Computational skill	Many correct	Few correct	.29
49.	M.I. 273	Fluency on self (absolute, not relative)	Many items	Few items	.46
50.	M.I. 161a	No. correct performances under approval relative to disapproval (comp. of lines)	High approval	High disapproval	.06
51.	M.I. 343	Rapid performance on approval relative to disapproval (formboard)	High disapproval	High approval	.74
52.	M.I. 343a	Faster performance under approval relative to disapproval (Pencil mazes)	High disapproval	High approval	.32
53.	M.I. 341	Higher accuracy to speed under approval vs. disapproval (mazes)	High disapproval	High approval	.20

TABLE 1. LIST OF OBJECTIVE TEST VARIABLES (continued)

No. of Variable in Factor Matrix	Master Index Number	Main Variables Title to be given to Positive Loading	High Test Score	Low Test Score	Reliability
54.	M.I. 205	Emotionality of comment	Many emotional comments	Few emotional comments	.08
55.	M.I. 97:123	Ratio passing time estimate, working in waiting	High waiting	High working	.63
56.	M.I. 212	Distant goals considered attainable	Few	Many	.67
57.	M.I. 268	Tempo of tapping	Many taps	Few taps	.65
58.	M.I. 31	Fluctuation of attitudes	Many changes	Few changes	.44
59.	M.I. 344	Tendency to like people	Many likes	Few likes	.19
60.	M.I. 345	Prolonged will-endurance (momentum)	Long time	Short time	.93
61.	M.I. 346	Will-endurance (increase in concentration)	Increase in time from I - II	Decrease in time from I - II	.41
62.	M.I. 347	Readiness to imitate	Many imitations	Few imitations	.79
63.	M.I. 348	Critical rejection-vs-uncritical acceptance of generalizations	Many rejections	Few rejections	.26
64.	M.I. 266	Ratio of initial to final performance	Much improvement	Little improvement	.26
65.	M.I. 101	Absolute level of aspiration	High	Low	.06
66.	M.I. 349	Consistency of aspiration	High & consistent	Low	.61
		Secondary Variables			
67.	M.I. 176:263:40	Slowness (vs. Instantaneousness) in false reaction times	Long for few	Short for Many	.44
68.	M.I. 72	Scatter in complex reaction times	High standard deviation	Low standard deviation	.69
69.	M.I. 350	Total errors in shape comparisons	Many	Few	.66
70.	M.I. 351	Shape comparison performance (errors and speed)	Many errors	Few errors	
71.	M.I. 353	Standard deviation in tempo of arm-shoulder movement	High	Low	.09
72.	M.I. 352	Standard deviation (unevenness in) number of objects seen in unstructured drawings	High	Low	.27
73.	M.I. 354	Possible responses missed on C.M.S. (Child version)	Many unmarked	Few unmarked	.57
74.	M.I. 355	Caution ("Don't know" responses) in Gestalt completion	Many "don't know"	Few "don't know"	.44
75.	M.I. 67	Memory for consonant statements (absolute no.)	High consonant	Low consonant	.31
76.	M.I. 356	Errors in color naming	Many	Few	.67
77.	M.I. 356a	Ideomotor speed: correct only	Many correct	Few correct	.82
78.	M.I. 2	Rigidity: color naming (corrected for speed)	Many correct	Few correct	.31
79.	M.I. 357	Venturesomeness in mirror drawing	Long distance covered	Short distance	.31
80.	M.I. 359	Annoyability (extremity of response)	Very annoyed	Little annoyed	.68
81.	M.I. 358	High performance on comparison of lines	Many correct	Few correct	.88

CONDITIONS OF THE EXPERIMENT

Due to the excellent and understanding cooperation of the University of Chicago Nursery School, for which we particularly wish to thank Mrs. Frances Prindle and her staff, it was possible to obtain four hours (eight half-hours individual sessions within a couple of weeks) of testing with each of 80 children. The choice of children was kept to as narrow an age range as possible, since we did not wish the

factor picture to be complicated through different maturation factors being created by the age variance. Both boys and girls were taken, 36 of the former, 46 of the latter, since we wanted to get factors general to the sexes. The children averaged 4.7 years of age, with a sigma of 0.4. They were somewhat selected for socio-economic status of parents, and, according to records, averaged well above the mean intelligence of the general community. However, they were mixed to the extent that there was a proportion of colored children among them, and, in personality, there seems no reason to suppose that any radical selection, of a kind that would completely obliterate a factor, can arise through social status or intelligence selection of a moderate kind.

Two experimenters carried out all the testing, about half the children being tested by each. They made themselves acquainted to the children before the test period, being several weeks on the premises, and the children were in all cases happy to go with them to the test room where they were put through the series of tests under the description of "games". The total test session was divided up for each child into eight parts, commonly occurring on different days, (but not remote, to avoid function fluctuation effects) with each part of about 30 minute duration. All of the objective tests were given in an individual test situation and in the same room. Wherever possible the tests were so arranged that each score could be split into two halves, similar to a split-half consistency coefficient, so that their reliabilities could be determined and allowed for in interpreting the subsequent factor structure.

In discussing these reliabilities it must be realized that we have deliberately entered this experiment with very brief tests (as we did throughout the whole research program) in order to find out the more promising tests by an initial broad "blocking in" of the factors. This permitted us to get 43 different tests, permitting 81 different behavioral scores into the available 4 hour total testing, by allowing only an average of 5 minutes per test. It is proposed hereafter to carry out, as we have done with the adult group, an improvement of the reliability of those tests having encouraging factor validity, by item analysis and lengthening of the test. However, for purposes of finding factor structure, reduction of loading by moderately low reliabilities is unimportant. Such reduced reliabilities lower the loadings on most factors, "across the board", but still reveal the essential factor patterns—the real desideratum at this stage.

The Factor Analytic Procedure

The test scores were intercorrelated by the product-moment formula and the 81 by 81 questionnaire matrix was then factor analysed by the centroid method. Tucker's criterion [4] and the new criterion worked out by Scheier, were tried in an attempt to decide exactly how many factors should be extracted. They pointed respectively to 17 and 19 factors. So, following the usual principle of extracting one or two factors in excess of the indicated number, and rotating them out thereafter, we enter the rotation process with 20 factors.

It is well known that factor analyses that are otherwise entirely sound and impeccable in procedure may become completely worthless through inadequately carried out rotation to simple structure [4]. Accordingly, a great deal of time was spent here in searching for the best possible structure, and in testing its significance when found. After the initial analytical oblimax rotation,[2] thirteen further, overall visual, single plane, scope plot rotations were carried out, reaching at one point 67% of variables in the + .10 hyperplane, but dropping, with adjustment of angles, to

[2]The oblimax analytical rotation method, developed by this laboratory in collaboration with K. Dickman and D. R. Saunders, maximizes the fourth powers, accepting oblique positions (whereof the orthogonal position is a permitted special case). However, experience has shown that both quartimax and oblimax do *not* rotate adequately to simple structure, though the latter is better. The analytical methods bring the rotation to a point where one can more readily see, in visual plots, what hyperplanes exist and what moves are necessary to converge upon them, but oblimax initially leaves the factors in a too oblique relationship, due to a few very highly loaded variables pulling the reference vector away from the perpendicular to its hyperplane.

59%. This gives a significance of simple structure, by Bargmann's index (1), of between 5% and 1%, but the significances are discussed afresh for individual factors below.[3]

Description of the Source Trait Findings

As usual, we shall set out the salients for each factor, from the rotated simple structure matrix (A. D. I. No. 5707) running down typically to correlations of 0.2, which, although well clear of the hyperplane, may be taken as the lower limit of adequate significance. However, where it is necessary to see what has happened to supposed markers for the factors in question, lower loadings will sometimes be taken. The marker variables for the factors hypothesized to be present will be written in italics. Factors are numbered and presented in order of declining contribution to the total variance of the 81 variables here studied.

Table 2. Factor 1: General Intelligence*

Variable	Description	Loading
70	Correct and valid shape comparison decisions	−62
69	Few errors in shape comparison decisions	−61
27	Low tendency to agree	−40
81	Good performance on comparison of lines	+39
24	High index of carefulness	+37
42	Quick learning of new names	+37
4	High ratio of irregular to regular warned reaction time	+35
26	Little critical exactness in correcting account of own performance	+32
34	High ideomotor speed (color naming)	+32
25	Good analysis on Gottschaldt figures	+26

*Note: This factor has been reversed relative to the matrix, for simplicity of interpretation.

This factor loads all the tests which would be expected to have Spearman's "g" in them, and also has the pervasive, large variance that one would expect of general intelligence in a battery of fairly complex tests applied to very young children. Its identification is confirmed by a correlation of −.8 (the factor happened to emerge here in reversed polarity) with the Children's Binet mental ages, a correlation not approached by any other factor. We set it aside here before proceeding with the personality factors proper.

Table 3. Factor 2: Willed-Responsiveness U.I. 29

Variable	Description	Loading
2	Fast irregularly warned reaction time	−66
8	High ratio of complex to simple reaction time	+58
1	Fast regularly warned reaction time	−57
18	Low ratio in correct to total responses on CMS performance (Child version)	−48
17	No excessive improper use of circles and slanting lines (equivalent) on CMS (Child version)	−41
19	*High total (correct) score on CMS (Child version)*	*+40*
*(53	Decline of accuracy to speed ratio, under approval, relative to disapproval	−39)
23	High speed of perceptual Gestalt completion	+34
5	Low variance of simple reaction time	−26
65	Low absolute level of aspiration	−25
**77*	*High ideomotor speed in color naming correct only*	+19
**34*	*High ideomotor speed (color naming)*	+17
* *3*	*Many reactions to false R.T. signal*	+16
**35*	*Low rigidity (reversed color naming)*	*−12*

*Reported for completeness; but of low reliability.
Markers for U.I. 29 in italics.

[3]The rotated reference vector matrix, V13, and the angles among the reference vectors, C13, are available, along with the unrotated factor matrix V0 and the transformation matrix T13, on microfilm from ADI under the number 5707. Send $1.25 for photostat or $1.25 for microfilm.

This pattern is comparable with U. I. 22 and U. I. 29 and no other, in the known adult and 11 year old series. In spite of the very high loading in simple reaction time, which normally characterizes U. I. 22, it is identified as U. I. 29, because the rest of the pattern of U. I. 29 is here, and because the next factor fits U. I. 22, though with somewhat lower absolute loadings on reaction time.

Variables 19, 3, 35, 34, and 77 have salient loadings, *i.e.*, loadings out of the hyperplane, here as required by the U. I. 29, *over-responsiveness* pattern. Speed of talking, present in the adult U. I. 29, fails here, while 4 and 50 fall in the hyperplane for the obvious reason that they have almost zero reliability. The general character of "keyed up" behavior persists in the new variables, but here one might question "*over-responsiveness*" since the performances are almost all in the direction of efficiency. Along with the GSR and metabolism associations in the adult patterns[6] one gets still, however, a concept of eagerness and capacity to mobilize rapidly one's resources. That is almost certainly a motivational source trait, most probably, as proposed elsewhere[6], connected with over-compensatory behavior for inferiority or insecurity.

TABLE 4. FACTOR 3: CORTICALERTIA U.I. 22

Variable	Description	Loading
68	Low variability in complex reaction time	−68
6	Fast complex reaction time	−56
47	High speed of repetitive formboard performance (realistic manipulation)	+56
5	*Low variance of simple reaction time*	*−55*
39	Low ratio of inaccuracy to speed (pencil maze)	−53
49	High fluency on self (absolute, not relative)	+46
1	*Fast regularly warned reaction time*	*−39*
2	*Fast irregularly warned reaction time*	*−36*
7	Many errors in maintaining complex reaction time sets	+34
8	*Many reactions to false R.T. signals*	*+31*
51	Rapid performance on approval relative to disapproval (formboard)	+31
34	*High ideomotor speed (color naming)*	*+19*
11	*Many objects seen in unstructured drawings*	*+10*
60	*Low dynamic momentum (will endurance)*	*−08*

All the markers for U. I. 22 appear here (and nowhere else except, in very incomplete form, in Factor 2, above) with correct signs and mainly with substantial loadings. It continues at the 4-5 year level to have the quality defined in adults as Corticalertia[6], namely, a high basic neural speed, with low variability, but tendency to "go off at half cock". The riddle of its distinction from U. I. 29 remains. Both are largely environmentally determined patterns, as to level[6], and both show a dynamic quality of eagerness and response to social approval, etc. The hyperplane fit falls below the P = .05 level here, but this is characteristically a pervasive factor.

TABLE 5. FACTOR 4: REALISM VS. SUBJECTIVITY U.I. 25

Variables	Description	Loading
34	*High ideomotor speed (color naming)*	*+81*
77	*High ideomotor speed in color naming (correct only)*	*+77*
35	High rigidity (reversed color naming)	+39
49	Low fluency on self (absolute, not relative)	−29
33	High fluency	+28
30	High willingness to forego prize for those in need	+28
46	Low speed of drawing	−26
58	Low fluctuation of attitude	−23
14	Low speed of social judgment	−23
31	Low extremity of viewpoint	−20
73	Few "decisions" (child equivalent) missed on CMS	−20
48	*High computational skill (Child version: comparison of quantities)*	*+17*
22	Low accuracy of perceptual Gestalt completion	−12
55	*Low ratio passing time estimate, occupied in idle periods*	*−10*

This is the only factor in the series with definite marker resemblance to U. I. 25, Realism, though it holds only for 34, 48, 22, and 55, the other required salients, 23, 45, and 57, falling in the hyperplane. The curious mixture of high speed on realistic, exact performances, *e.g.*, computation, with rigidity in more creative or, at least, subjective response, *e.g.*, drawing, fluency on self, gestalt completion, social judgement, low attitude fluctuation, remains, and agrees with the Realism-*vs.*-Subjectivity (Eysenck "psychoticism") interpretation of U. I. 25[6]. However, both this factor meaning and the hyperplane are not well defined.

TABLE 6. FACTOR 5: ANXIETY U.I. 24

Variables	Description	Loading
43	*High annoyability (all situations)*	+79
80	Many "extreme" annoyance responses	+76
28	*Willingness to admit being naughty (common frailties)*	+28
47	Low speed of repetitive formboard performance (realistic manipulation)	−27
31	Low extremity of viewpoint	−26
6	Slow complex decision reaction time	+26
9	Low speed of decision in visual comparisons	−23
42	High readiness to learn new names	+20
54	*High emotionality of comment*	*+16*
58	*Low fluctuation of attitudes*	−14
27	*High tendency to agree*	*+02*

The now well established Anxiety, U. I. 24 pattern[6, 14], has resemblance initially here to factors 5 and 14, but there are good reasons for preferring the former, despite variable 16 loading in the wrong direction (discountable in view of its essentially zero reliability). All remaining markers have the right sign, though some are of very small magnitude compared with the adult pattern, and indeed "Tendency to Agree" virtually fails. Further experiment might be done with rotations of this factor. Variable 42, High Readiness to Learn, perhaps is the forerunner of the Good Ethical Choice and Absence of Questionable Reading Preferences which load the factor in later life. An objective anxiety battery for pre-school children has been added, by present results, to those available at higher ages[14].

TABLE 7. FACTOR 6: NEURAL RESERVES-VS.-NEUROTICISM U.I. 23

Variable	Description	Loading
78	High liability to error (relative speed) in "creative effort" (Rigidity function)	+63
76	Many errors in rapid color naming	+45
65	*Low absolute level of aspiration*	*−44*
38	*Low preference for color over form in pictures*	*−41*
72	High variability in number of objects seen in unstructured drawings	+37
35	*Low rigidity (reversed color naming)*	*−32*
39	*Low ratio of inaccuracy to speed (pencil maze)*	*−29*
47	Low speed on repetitive formboard performance (realistic manipulation)	−27
46	Low speed of drawing	−26
51	Slow performance on approval relative to disapproval (formboard)	−25
60	*High will-endurance (second trial)*	+14
48	*High computational skill (Child version: comparison of quantities)*	*+12*
33	*High fluency*	*+09*

Factors with any resemblance to Neural Reserves-*vs*-.Neuroticism, U. I. 23, are 10, 19, and the present pattern; but this alone clearly has the main markers[6]. However, "Low absolute level of aspiration" must be discounted because of its low reliability, and "High fluency", though on the right side of the hyperplane, is negligibly loaded. It is interesting that the child scoring highest in the non-neurotic direction is most prone to careless errors in color naming and to low aspiration. The hyperplane is at the P = .05 level.

TABLE 8. FACTOR 7: HARRIC ASSERTIVENESS: U.I. 16

Variable	Description	Loading
65	High absolute level of aspiration	+56
56	Many distant goals considered attainable	−47
66	High consistency of aspiration	+37
41	*Much acceptance of obvious moral aphorisms (platitudes)*	*+35*
27	Low tendency to agree	−33
61	Increased will-endurance (second trial)	+27
70	Good shape comparison performance	−25
26	Little critical exactness in correcting account of own performance	−24
57	*Fast tempo of tapping*	*+23*
6	Fast complex reaction time	−22
59	Tendency to like people	+22
25	Good analysis performance on Gottschaldt figures	+21
9	*High speed of decision in visual comparisons*	*+20*
20	Low ratio self criticism to criticism of others	−20
39	High ratio of inaccuracy to speed (pencil maze)	+20
44	*Good ethical choices*	*+18*

Possible rival pattern matches for U. I. 16 were factors 4 and 6. It is noteworthy that this largest variance personality factor, U. I. 16 in adults, is large here also. A few possible adult markers: 11, 17, 19, 35, 77, and 81—mainly concerned with willed speed (and contact)—fail to load here, and it is brought out more clearly than in the adult that the factor is concerned more with tempo than with speed. However, there is the same central hypothesized character[6, p. 236] of "expressiveness and assertiveness, with fast tempo," here showing in high aspiration, will endurance, little tendency to agree, and both kinds of tempo. There are also associations fitting the correlations found[6] with the 16 PF questionnaire factor of Harria. Finally there is the usual emphasis[6, p. 235] on good social values. (Variables 41 and 44).

TABLE 9. FACTOR 8: STOLIDITY, COOLNESS IN STRESS: U.I. 30

Variable	Description	Loading
18	Low ratio incorrect (forbidden) to total responses on CMS performance (Child version)	−51
17	No excessive use of circles and slanting lines (equivalent) on CMS (Child version)	−47
11	Few objects seen in unstructured drawings	+43
19	High total (correct) score on CMS (Child version)	+40
68	Low variability in complex reaction times	−40
60	Absence of will-endurance (dynamic momentum, physical stress)	−39
12	Few threatening objects in unstructured drawings	−37
62	Low readiness to mimic	−28
10	*Slow tempo of arm-shoulder movement*	*−21*
58	*Low fluctuation of attitude*	*−13*

This is the only factor containing the markers for U. I. 30 with the right sign combination; but its identification must be tentative because it also contains variables, notably the C. M. S., not previously found to load U. I. 30. Its sense fits the "stolidity" title assigned to U. I. 30[6] and suggests steadiness, and coolness in stressful situations.

TABLE 10. FACTOR 9: AUTIA-VS.-PRAXERNIA: U.I. 34

Variable	Description	Loading
9	High speed of decision in visual comparisons	+44
6	Fast complex decision reaction time	−43
12	*Many (absolute no.) threatening objects in unstructured drawings*	*+42*
38	Low preference for color over form in pictures	−41
8	Low ratio of complex to simple reaction time	−39
11	*Many objects seen in unstructured drawings*	*+36*

TABLE 10. FACTOR 9: AUTIA-VS.-PRAXERNIA: U. I. 30 (continued)

Variable	Description	Loading
30	High willingness to forego prize for those in need	+33
33	High fluency	+33
70	Good shape comparison decisions	−28
21	Much criticism of performances (own and others)	+25
80	*Many "extreme" annoyance responses*	*+24*
16	*Many questionable preferences in reading*	*+15*
7	*Many errors of decision in maintaining complex reaction time sets*	*+10*

This is a very good match for the U. I. 34 pattern which Rosenthal located [19], and Scheier confirmed [14], since the original survey of replicated factors [6]. The character of the present new, non-marker variables is, moreover, fully consistent with the identification as M, autia, [6], made by Scheier [14], namely, imaginative, fluent, quick, but fearful, easily critical and introverted. It is increasingly apparent that M, autia, despite its late discovery, is a personality factor of considerable stability of form, and practical importance.

TABLE 11. FACTOR 10: CRITICAL RESTRAINT (NEW). POSSIBLY U.I. 26

Variable	Description	Loading
22	High accuracy of perceptual Gestalt completion	+55
74	High caution ("Don't Know" responses) in Gestalt completion	+55
20	High ratio self criticism to criticism of others	+38
(50	Few correct performances under approval relative to disapproval (comp. of lines)	−32)
61	Increased will-endurance (second trial)	+31
54	High emotionality of comment	+28
5	Low variance of simple reaction time	−24
69	Few errors in shape comparison decisions	−24
59	Tendency to dislike people (in portraits)	−23
7	Few errors of decision in maintaining complex reaction time sets	−23

This has some resemblance to U. I. 17, 23, and 29, but most to U. I. 26. However, the agreement is not good enough to conclude any definite match and it is tentatively called, descriptively, *Critical Restraint*, and left to future verification as a pattern (incidentally, with a better hyperplane than here) before theorizing. This is at least highly consistent in its variables, depicting strong will, exactitude, criticism of self and others, and cautiousness. Its resemblance to the *sense* of U. I. 17 is great, yet the markers for U. I. 17 fall in Factor 12, so we must hypothesize two sources of inhibition.

TABLE 12. FACTOR 11: CRITICAL PRACTICALITY: U.I. 19

Variable	Description	Loading
13	Low percentage threatening objects perceived in unstructured drawings	−51
7	*Few errors of decision in maintaining complex reaction time sets*	*−44*
22	*High accuracy of perceptual Gestalt completion*	*+42*
31	High extremity of viewpoint	+42
3	Few reactions to false R.T. signal	−37
76	Few errors in rapid color naming	−37
24	*High index of carefulness (Picture recognition)*	*+37*
68	High variability in complex reaction times	+30
69	Few errors in shape comparison decisions	−27
79	Low trial and error activity in mirror drawing	−27
25	*Good analytic performance on Gottschaldt figures*	+18
20	*Low ratio self criticism to critism of others*	*−14*

This is an excellent match for adult and 11-year old Critical Practicability, U. I. 19, having all definite markers equated and even less exact markers, *e.g.*, low

tendency to imitate, few slanting lines, in CMS, and low rigidity, correct in sign. The new variables bear out the character of accuracy and penetration, with outward direction of aggression.

TABLE 13. FACTOR 12: INHIBITION: U.I. 17

Variable	Description	Loading
15	*Willingness to play practical jokes and tease*	*+40*
72	High variability in no. of objects seen in unstructured drawings	+36
23	*Slow speed of perceptual Gestalt completion*	*−36*
38	Preference for form over color in picture	−35
5	Low variance of simple reaction time	−35
71	Low variability in tempo of arm-shoulder movement	−35
11	Many objects seen in unstructured drawings	+34
12	Many (absolute no.) threatening objects in unstructured drawings	+31
21	Little criticism of performances (own and others)	−30
16	*Few questionable preferences in reading*	*−28*
14	*Slow speed of social judgment*	*−16*
41	*Much acceptance of obvious moral aphorisms (platitudes)*	*+10*
13	*High percentage threatening objects perceived in unstructured drawings*	*+10*

This and this alone, adequately carries the markers for U. I. 17 though, as usual, a slight cooperativeness is seen with U. I. 26 (Factor 19 here). Only the slowing of reaction time with complex instructions fails to load sufficiently, though the difference is in the right direction. The new variables mainly bring in a low variability in speed and tempo. However, the emphasis has changed from that among adults. Thus variable 15 which hypothetically indicates an outlet for the highly inhibited (see Oliver Cromwell) comes higher, while 13 (as found in 11-year-olds) comes lower in the pattern [6]. The hyperplane significance is at P = .01.

TABLE 14. FACTOR 13: MIRROR-DRAWING DOUBLET

Variable	Description	Loading
37	Slow and defective performance on mirror drawing	+70
36	Poor final achievement (orientation rigidity) in mirror drawing	+62
61	Increased will-endurance (second trial)	+28
16	Few questionable preferences in reading	−26
54	Low emotionality of comment	−24

No match can be found for this, which is scarcely surprising, seeing that most of its variance arises artifactually, as a doubtlet from two ways of scoring the mirror drawing behavior, but its hyperplane is sound (P = .01).

TABLE 15. FACTOR 14: COMENTION-VS.-ABCULTION: U.I. 20

Variable	Description	Loading
63	Readiness critically to reject loose generalizations	+67
27	*High tendency to agree*	*+47*
28	*Willingness to admit being naughty (common frailties)*	*+25*
25	Good analytic performance in Gottschaldt figures	+24
60	High will-endurance (second trial)	+24
15	Not willing to play practical jokes and tease	−23
21	Much criticism of performances (own and others)	+23
10	Slow tempo of arm-shoulder movement	−23
11	Many objects seen in unstructured drawings	+23
29	Tendency to recall consonant-vs.-disonant statements	+11

With the highest of all loadings on 27 and 28, the two highest markers for U. I. 20, (Maturity of Opinion had zero reliability here) there can be little doubt in identi-

fying this as U. I. 20, though the hyperplane is relatively poor (52%). Yet there exists a not previously noted cooperativeness with U. I. 30 (slight and on two variables) and resemblance thereto in terms of "stolidity" and "independence." The main meaning remains in accordance with our hypothesis of a general cooperativeness and tendency to accept, and impose on the self, good cultural standards[6].

TABLE 16. FACTOR 15: EXUBERANCE: U.I. 21

Variable	Description	Loading
14	*High speed of social judgment*	*+49*
29	Tendency to recall dissonant rather than consonant statements	−48
23	High speed of perceptual Gestalt completion	+46
33	*High fluency*	*+31*
9	*High speed of decision in visual comparisons*	*+30*
69	Many errors in shape comparison decisions	+28
75	*Many agreed responses remembered*	*+25*
57	*Fast tempo of tapping*	*+24*
62	High readiness to mimic	+21
80	Many "extreme" annoyance responses	+21
54	*High ideomotor speed (color naming)*	*+18*
27	*High tendency to agree*	*+12*

Unfortunately, the highest marker for U. I. 21 (No. 50 here) had to be rejected (Table 1) for low reliability; but otherwise the markers for U. I. 21 appear very consistently here (and nowhere else, except for slight resemblance to Factor 4). (Computational skill is positive, but falls into the hyperplane. Its failure is presumably our failure to construct a 4 year old equivalent for computation.) The descriptive hypotheses of "exuberance" is continued[6, p. 148] in the new variables, *e.g.*, many errors, readiness to mimic, extremity in statements of annoyance.

TABLE 17. FACTOR 16: EXAVIA-VS.-INVIA: U.I. 32

Variable	Description	Loading
48	High computational skill (Child version: comparison of quantities)	+45
62	High readiness to mimic	+37
67	Tendency of false reactions to occur quickly	−36
66	Little consistency of aspiration	−32
52	Slow performance on approval relative to disapproval (pencil maze)	−31
49	*High fluency on self (obsolute, not relative)*	*+25*
31	Low extremity of view point	−23
81	Accurate and fast performance on comparison of lengths	+23
74	*Low caution ("Don't Know" responses) in Gestalt completion*	*−22*
4	High ratio of irregular to regular warned reaction time	+21
20	Low ratio self criticism to criticism of others	−21
11	Many objects seen in unstructured drawings	+11
17	*No subjective, excessive use of circles and slanting lines (equivalent) on CMS test (Child version)*	*−10*
22	*Low accuracy of perceptual Gestalt completion*	*−04*

Four markers existed in this experiment for U. I. 32, now firmly identified as what is commonly called extraversion-*vs.*-introversion, a second order rating and questionnaire factor,[6, pp. 166, 317]. They fall here consistently but with low loadings, as shown, in a factor which has rather surprisingly small variance. Its hyperplane is also poor. Presumably, the large change in age has brought a reduced projection, while recognizably retaining the pattern, and the measurement of the exavia factor will need to be transferred here to new variables. The new variables, however, fit very well the notion of attention to external reality, readiness to mimic, lack of caution, and self criticism, and impulsiveness and lack of consistency, in extraversion.

TABLE 18. FACTOR 17: U.I. 35

Variable	Description	Loading
55	Low ratio passing time estimate occupied in idle periods	−48
58	Low fluctuation of attitude	−39
48	Low computational skill (Child version: comparison of quantities)	−33
40	Low increase in accuracy (to speed) on passing from easy to difficulty maze	−33
81	Accurate and fast performance on comparison of lines	+30
26	Little critical exactness in correcting account of our performance	−28
73	Many "decisions" (child equivalent) missed on CMS	+28
36	Poor final achievement (orientation rigidity) in mirror drawing	+25
4	High ratio of irregular to regular warned reaction time	+19)
43	High annoyability (all situations)	+15
49	High fluency on self (absolute, not relative)	+10

The pattern of U. I. 35 has only been recognized and replicated, in adults, recently by Scheier. Three of the four markers can be accepted as replicating the pattern here in children, with reduced projection however, and therefore permitting us to indicate this only as a tentative pattern in children, awaiting confirmation in terms of the better variables here found.

TABLE 19: FACTOR 18: ASTHENIA (WITH GOOD COMPLEX PERFORMANCE): U.I. 28

Variable	Description	Loading
75	Many agreed responses remembered (Consonant statements)	+46
51	Slow performance on approval relative to disapproval (formboard)	−33
3	Many reactions to false R. T. signals	+32
6	Fast complex decision reaction time	−31
17	Excessive use of circles and slanting lines (equivalent on CMS test (Child version)	+31
9	Low speed of decision in visual comparisons	−28
62	Low readiness to mimic	−26
8	Low ratio of complex to simple reaction time	−22
46	High speed of drawing	+22
57	Slow tempo of tapping	−21
55	High ratio passing time estimate occupied in idle periods	+17
56	Few distant goals considered attainable	+13
49	High fluency on self (absolute, not relative)	+13
43	High annoyability (all situations)	+11
27	High tendency to agree	+10
29	Tendency to recall consonant vs. disonant statements	+03

The markers, in descending order 55, 27, 43, 49, 29 for U. I. 28 appear here, and only here, simultaneously with loadings outside the hyperplane. Although small, all five are correct in sign, and if 29 is combined with 75, as might be appropriate, the cummulate significance is high. However, U. I. 28 remains here, as in other studies, a factor of low variance, poorer hyperplane and elusive meaning. (The interpretation favoring a combination of emotional impulsiveness with asthenic qualities, of high speed on some complex performances with slow tempo, etc.) It deserves special investigation, in terms of finding one or two substantially loaded markers to aid interpretation.

TABLE 20. FACTOR 19: SELF SENTIMENT (CONTROL): U.I. 26

Variable	Description	Loading
64	*Low ratio of initial to final (simple) performance*	*+57*
35	*Low rigidity (reversed color naming)*	*−55*
78	Lower liability to error (rel. speed) in "creative effort" (rigidity function)	−37
42	Readiness to learn new names	+35
34	Low ideomotor speed (color naming)	−28
77	Low ideomotor speed in color naming (correct only)	−24
18	*High percentage threatening objects perceived in unstructured drawings*	*+21*
37	Opposite: Slow and defective performance in mirror drawing	−21
60	*High will-endurance (second trial)*	*+13*

This can be considered only fairly good as a match for U. I. 26, because the highest marker in the latter "high ratio of fluency on self to other topics," was not available as a ratio here. The match rests, therefore, on 60, 35, 13, and 64 (that order in original[6]), while 7, 32, 33, 38, 50, 54, and 75, despite having only one exception to correct sign pattern, are never clear of the hyperplane. If the hypotheses[6] that this represents the self-sentiment is correct, one might expect rather large modification of the pattern between 4 and 11 years. The hyperplane fit is very good: $P < .001$.

TABLE 21. FACTOR 20: U.I. 18

Variable	Description	Loading
64	*Low ratio of initial to final (simple) performance*	*+50*
57	Slow tempo of tapping	−33
56	Few distant goals considered attainable	+28
15	Not willing to play practical jokes and tease	−26
19	*High total (correct) score on CMS (Child version)*	*+23*
35	Low rigidity (reversed color naming)	−23
74	Low caution ("Don't Know" responses) in Gestalt completion	−23
45	High expansiveness of copied design	+22
29	*High ratio of consonant to dissonant opinions recalled*	*+16*
17	*Excessive use of circles and slanting lines (equivalent) in CMS test (Child version)*	*+15*

This is the only possible match for U. I. 18, but although the markers occur together, No. 29 is reversed in direction. However, the hyperplane is very significant and the general sense (note 56, 35, 74) fits what, for lack of better terms has been called the hypomanic tendency in U. I. 18. And "expansiveness of copied design" may perhaps be counted a rough equivalent for the marker, not used in this study, "high speed on finger maze."

SUMMARY AND CONCLUSIONS

1. Eighty-one behavioral response measures on objective tests chosen to mark, with suitable modification, factors known in older subjects, and covering the general personality sphere, have been factored on eighty 4 or 5 year olds, yielding twenty factors revealing a simple structure significant at the $P < .05$ to $< .01$ level.

2. Omitting inferences from loadings on a dozen variables of low reliability the loading patterns in the remainder can be matched with older child and adult personality factor patterns U. I. 16 through 35, leaving only U. I. 27, 31 and 33 unmatched in the standard replicated personality factor series[6], and only numbers 10 and 13 unmatched in the present series. Number 13 is virtually only a doublet, but 10 is a new factor of definite psychological consistency, called Critical Restraint.

3. The matching is very definite in fourteen factors, in that: (a) all markers are present, most with non-hyperplane loading, and all with correct sign, (b) no truly competitive *alternative* match exists on *either* side, and (c) the new, non-marker variables fit the psychological sense of the factor as previously hypothesized.

However, in four others the matching can be considered only tentative, and in all there is a tendency for the marker pattern to appear with reduced variance, probably showing greater variance loss than in previous cross validations confined to groups of the same age as that on which the factors were first recognized. Also, the general rank order of variance magnitude of the factors is only slightly related to that in adults. Accordingly, we may tentatively conclude that there are systematic changes in the factor pattern emphasis between four years of age and adulthood, though the number and nature of the factors stays remarkably constant.

4. To check the identifications it is desirable that this study be independently repeated on a new 4-5 year old sample, but it would be valuable to do this while: (a) increasing the length of tests which proved deficient in reliability, (b) shifting the matching emphasis in determining relation to the present study to *new* variables

found loaded higher here than at upper ages, (c) examining the hypotheses by inserting new tests aimed at the essence of present factors, and, (d) observing the agreement of the present nursery school group modification from the adult-marker patterns with the modifications that appear in our study of 7 and 8 year olds, now in process of publication [9].

The factors checked here—but especially the general anxiety factor U. I. 24, the general neuroticism factor U. I. 23, and the extraversion-introversion factor, U. I. 32—should provide a useful basis of determinate, uniquely-defined measures for those concerned with studies of personality development in early childhood. A best possible battery for twelve factors has been put together from the tests in the present research, and will be made available to clinical researchers on request.

Bibliography

1. Bargmann, R. Signifikanzuntersuchungen der einfachen Struktur in der Factoren-analyse. Wurzburg: Sonderdruck, Physica Verlag, 1955.
2. Cattell, R. B. The description of personality: I. Foundations of trait measurement. *Psychol. Rev.*, 1943, *50*, 559-594.
3. Cattell, R. B. Interpretation of the twelve primary personality factors. *Char. and Pers.*, 1944, *13*, 55-92.
4. Cattell, R. B. *Factor Analysis*, New York: Harper Bros., 1952.
5. Cattell, R. B. *The objective-analytic personality factor battery*. I. P. A. T. 1602 Coronado Drive., Champaign, Illinois, 1955.
6. Cattell, R. B. *Personality and Motivation Structure and Measurement*. New York: World Book Co., 1957.
7. Cattell, R. B. What is "objective" in "objective personality tests?" *J. consult. Psychol.*, 1958, vol. 5, 4, 285-289.
8. Cattell, R. B. and Coan, R. W. Child personality structure as revealed in teacher's behavior ratings. *J. clin. Psychol.*, 1957, *13*, 315-327.
9. Cattell, R. B. and Coan, R. W. Objective test assessment of the primary personality dimensions in middle childhood. *Brit. J. Psychol.* (In Press).
10. Cattell, R. B. and Coan, R. W. Personality dimensions in the questionnaire responses of six and seven-year-olds. *Brit. J. Educ. Psychol.*, 1958, *28*, 232-242.
11. Cattell, R. B. and Gruen, W. Primary personality factors in the questionnaire medium for children eleven to fourteen years old. *Educ. Psychol. Measmt.*, 1954, *14*, 50-76.
12. Cattell, R. B. and Gruen, W. The primary personality factors in eleven year old children by objective tests. *J. Personal.*, 1955, *32*, 460-478.
13. Cattell, R. B. and Peterson, D. R. Personality structure in 4-5 year olds, by factoring observed, time sampled behavior. *Ras. Psico, Gen. Clin.*, 1958, *3*, 1-21.
14. Cattell, R. B. and Scheier, I. *The Meaning and Measurement of Neuroticism and Anxiety*. The Ronald Press., (In Press).
15. Koch, H. L. A multi-factor analysis of certain measures of activeness in nursery school children. *J. genet. Psycho'.*, 1934, *45*, 482-487.
16. Peterson, D. R. and Cattell, R. B. Personality sturcture in 4-5 year olds by factoring oral questionnaire responses. (In Press).
17. Peterson, D. R. and Cattell, R. B. Personality structure in the nursery school child as revealed by parents' ratings. *J. clin. Psychol.* (In Press).
18. Peterson, D. R. and Cattell, R. B. Personality factors in nursery school children as derived from parent ratings. *J. clin. Psychol.*, 1958, *14*, 346-355.
19. Rosenthal, I. A factor analysis of anxiety variables. Ph.D. Thesis, University of Illinois Library, Urbana, 1955.

THE MEANING OF A POSITIVE SELF-CONCEPT*

FREDERICK M. JERVIS
University of New Hampshire

PROBLEM

There is general agreement, both in self-concept theory and in experimental evidence, that a negative self-concept indicates stress and tension within the individual. In contrast, interpretations of the significance of a positive self-concept have varied. A large body of research, identified with Carl R. Rogers and the Chicago School, was based on the assumption that a positive self-concept indicated a well-adjusted individual. Both in theory and research design, this group stressed the importance of self-integration with the implicit assumption that the well-integrated individual would also be integrated into his society.

Rogers and Dymond [11] have recently suggested that the positive self-concept may also indicate a highly defensive paranoid individual. Frenkel-Brunswik [2] and Hart [5] offer further support of the theory that positive self-concept, for some individuals, may be compensatory in nature. Hart recognized the difficulty in differentiating between true self-esteem and inflated self-esteem. The possibility that high self-esteem can compensate for underlying negative attitudes toward self has important theoretical implications.

The purposes of this study are to examine the heterogeneity of the positive self-concept group, and to determine whether homogeneous sub-groups are related to external criteria. A positive self-concept is defined as a small discrepancy between self and ideal-self ratings. It was hypothesized that quantitative differences exist within a group of individuals expressing positive self-concept.

METHOD

Self-Concept. The experimenter constructed a Self Description Inventory (SDI), a measure of self-concept which is expressed as a discrepancy (D), and which provides in addition, self (S) and ideal-self (I) scores for each subject providing a basis of identifying quantitative differences within the positive self-concept group. The portion of the SDI used in this study consists of 60 items rated on a five point scale (a) as the individual describes himself and (b) as the individual would ideally like to be. The discrepancy between the two responses constitutes the score for each item; item discrepancies are totaled to yield a measure of self-concept (D). Reliabilities of the three SDI scales are shown in Table 1. Detailed information about item selection, derivation of self and ideal-self scores, and reliabilities of parts of the scales can be found in the author's doctoral dissertation [7].

TABLE 1. RELIABILITIES OF SDI SCALES

SDI Scale	N	Split-half	N	Test-retest*
Discrepancy	110	.93	215	.68
Self	100	.90	215	.83
Ideal-self	100	.90	215	.39

*Two months elapsed time.

Positive Self-Concept. The distribution of SDI discrepancy scores was divided into quarters. The highest quarter (D^4) was defined as the negative self-concept group. The lowest quarter (D^1) was designated the positive self-concept group and was treated as a discrete sample. Self scores of D^1 (positive self-concept) individuals were ranked and the distribution was divided into thirds. The lowest third of the self scores was designated s^1, the middle third was designated s^2, and the highest third was designated s^3. Individuals with positive self-concepts were also ranked according to ideal-self scores and discrepancy scores. These rankings were divided into thirds and designated similarly: i^1, i^2, i^3, and d^1, d^2, and d^3. Every individual with a positive self-concept was in one of the three categories on each of the scales.

*This article is based on a doctoral dissertation submitted to Teachers College, Columbia University. Members of the dissertation committee were Professors Laurence F. Shaffer, Chairman; Arthur T. Jersild; and Paul E. Eiserer.

Attitudes Toward Others. Theoretical concepts[8, 9, 10] as well as experimental findings[1, 9, 12, 13] support the thesis that positive attitude toward self is correlated with positive attitude toward others. Attitude toward others was therefore selected as an external criterion for differentiating among subgroups of individuals with positive self-concepts. The Chicago Self-Others Scale which measures attitude toward a "generalized" other was used in this study.

Predicted Academic Achievement. A possible criterion for differentiating individuals with positive self-concepts was the accuracy with which they predict future achievement. Subjects predicted their academic achievement on a seven point scale. In order to obtain a difference score between predicted and actual achievement, grade point averages of subjects were given a value according to their position on a seven point scale.

Academic Achievement. Academic achievement was selected as a criterion which might differentiate subgroups of individuals with positive self-concepts because of theoretical indications that highly defensive individuals may be characterized by low achievement[6, 14].

Subjects and Administration of Tests. The SDI was administered as part of the Orientation Week testing program to the 850 freshmen entering a state university in 1956. Before students had entered their first examination period, they were asked to predict their relative academic standing for the first semester (N = 761). The Chicago S-O Scale[4] was administered during the second semester of the subjects' freshman year (N = 563). Grade point averages were available in the Recorder's Office for 791 of the students who completed the fall semester.

Results

Interrelationships of SDI Scales. Intercorrelations between the three scales were: Discrepancy-Self $r = -.57$; Discrepancy - Ideal-self $r = .30$; and Self - Ideal-self $r = .37$.

Women's D scores were slightly higher than men's (significant at the .05 level of confidence). The ideal-self scores of the women were also higher than those of the men (significant at the .01 level of confidence). The higher discrepancy scores for women were accounted for by their higher ideal-self scores. See Table 2.

Table 2. Sex Differences on SDI Scales

Measure	Men (N = 556) M	SD	Women (N = 294) M	SD	t
Discrepancy	47.96	22.6	51.6	21.2	2.31*
Self	217.6	23.1	218.9	23.7	.83
Ideal-self	247.6	23.6	254.9	22.5	4.64**

*Significant at the .05 level.
**Significant at the .01 level.

The self scores of individuals with positive self-concept ranged from 173 to 271 with an M of 231 and SD of 20. The ideal-self scores of the same subjects ranged from 169-295 with an M of 239 and SD of 24. This group was characterized both by low self and low ideal-self ratings, as well as by high self and high ideal-self ratings.

Although the three SDI scales were significantly related, the coefficients of determination suggest that the scales measure separate but related aspects of an individual's self-concept. In addition, the heterogeneity of the positive self-concept group can best be demonstrated by the variety of self – ideal-self combinations which resulted in a small discrepancy score.

Attitude Toward Others. Because of the lapse of time between the administrations of the SDI and the S-O Scale, the SDI was readministered to a random sample of 98 students. No significant differences were found when the S-O Scale was related to the two administrations of the SDI. Therefore, all subjects who had taken both the SDI and the S-O Scale were used in the study. The discrepancy scale of the SDI and the S-O Scale correlated −.03. Attitude toward others was not related to self-concept. Analysis of S-O scores within D quarters showed in both D_1 (positive self-concept) and D_4 (negative self-concept) a significantly greater number of men scoring

in the highest S-O scale quarter (negative attitude toward others) than would be expected by chance. $X^2 = 12.5$ and 5.6 respectively, with both significant beyond the .05 level. Both positive and negative self-concept groups of men were characterized by negative attitudes toward others.

TABLE 3. MEANS AND STANDARD DEVIATIONS OF S-O SCORES WITHIN D^1 SUBGROUPS ($N = 151$)

D^1 Subgroups	N	M	SD
Self: s^3	46	392.11	42.74
s^2	49	413.66	49.16
s^1	56	416.87	47.62
Ideal: i^3	46	390.37	41.62
i^2	43	421.59	44.64
i^1	62	414.98	49.80
Discrepancy: d^3	49	402.97	46.66
d^2	50	417.90	53.54
d^1	52	407.96	41.46

Table 3 shows that s_1 and s_2 groups scored significantly higher on the S-O Scale than s_3 ($t = 2.75$ and 2.25 respectively). Similarly, i_1 and i_2 groups scored significantly higher on the S-O Scale than did i_3 ($t = 2.77$ and 3.37 respectively). When the d groups were analyzed on the basis of S-O scores, the t values were not significant.

A positive self-concept, as measured and defined in this study, was not always correlated with a positive attitude toward others. Low discrepancy scores characterized by low self scores and low ideal-self scores were also characterized by negative attitude toward others.

Predicting Academic Achievement. Results disclosed no significant relationship between self-concept and prediction of grades. Prediction of grades did not differentiate subgroups of individuals with positive self-concepts. The D_1 group differed from the other D groups in the direction of overestimating grades. Only in the comparison of the D_1 and D_2 groups was the difference significant ($t = 2.65$, significant at the .01 level). Results suggest only a tendency toward overestimation of academic achievement by positive self-concept individuals.

Actual Academic Achievement. There was no significant relationship between self-concept scores and grades. Academic achievement differentiated only women among defined subgroups of individuals with positive self-concepts. The s_1 group of women had significantly lower grades than the s_2 and s_3 women ($t = 3.43$ and 3.13 respectively). In addition, the grade point average of women in the s_1 group was significantly lower than the mean grade point average of freshman women ($t = 3.41$).

ACE scores of women with positive self-concepts were examined to determine whether differences in academic aptitude accounted for the low academic achievement of s_1 women. Q scores of the three s groups of D_1 women did not differ significantly. However, the L score was significantly lower for the s_1 group. It is possible that the self-concepts of these individuals have operated to inhibit the type of learning predicted by the Linguistic Scale of the ACE.

SUMMARY

1. The self-concepts of 850 college students were measured and related to attitude toward others, prediction of academic achievement, and actual academic performance.

2. Results disclosed that the positive self-concept group contained individuals with low self and low ideal-self scores as well as individuals with high self and high ideal-self scores.

3. Positive self-concept subgroups were differentiated by attitude toward others. Men and women responded differently to the self-concept scale and to the external criteria.

References

1. Berger, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, *47*, 778-782.
2. Frenkel-Brunswik, Else. Mechanisms of self deception. *J. soc. Psychol.*, 1939, *10*, 409-420.
3. Fromm, E. *Man for Himself*. New York: Rinehart, 1947.
4. Gordon, T. and Cartwright, D. S. The effect of psychotherapy upon certain attitudes toward others. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
5. Hart, H. H. Narcissistic equilibrium. *Int. J. Psycho-Anal.*, 1947, *28*, 113.
6. Horney, Karen. *New ways in psychoanalysis*. New York: Norton, 1939.
7. Jervis, F. M. The meaning of a positive self-concept. Unpublished doctoral dissertation, Teachers College, Columbia Univer., 1948.
8. Lecky, P. *Self-consistency*. New York: Island Press, 1945.
9. Phillips, E. L. Attitudes toward self and others: a brief questionnaire report. *J. consult. Psychol.*, 1951, *15*, 79-81.
10. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
11. Rogers, C. R. and Dymond, Rosalind F. *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
12. Scheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, *13*, 169-175.
13. Stock, Dorothy. An investigation into the interrelations between the self-concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, *13*, 176-180.
14. Symonds, P. M. *The ego and the self*. New York: Appleton-Century-Crofts, 1951.

DIFFERENCE BETWEEN NEUROTICS, PSYCHOTICS AND NORMALS ON PERCEPTUAL TESTS: REANALYSIS OF EYSENCK'S RESULTS

STEVEN G. VANDENBERG*

University of Michigan

Problem

Recently Eysenck(1) reported the results of a comparison of the performances of 20 neurotics, 20 psychotics, and 106 normal males on a variety of objective tests yielding a total of 91 scores. Included in this report are tables showing means and variances for the three groups plus one F ratio of the between-groups and the within-groups variances. This statistical technique capitalizes on three comparisons: Neurotics *vs.* psychotics, neurotics *vs.* normals, and psychotics *vs.* normals. It seems of interest to know the results of each of these three comparisons separately. Eysenck states that neuroses are not merely milder forms of mental illness than the psychoses. He contends that neuroticism and psychoticism form two orthogonal dimensions of personality. If this view is correct, we would expect that the means of the three groups would not always be in the order: normal, neurotic, psychotic.

Method

The *t*-test was used to determine the statistical significance of the difference between the means of the neurotics (Neur.) and the psychotics (P) on each of the

*The analysis reported here was supported by the Schizophrenia and Psychopharmacology Study, a Joint Research Project of the University of Michigan and Ypsilanti State Hospital, Ralph W. Gerard, Principal Investigator.

tests, as well as between the means of the neurotics and of the normals (Nor.), and between the means of the psychotics and of the normals.

Results

Table 1. The Mean of the Neurotics Lies Between the Mean of the Normals and the Mean of the Psychotics. A Reversed Order is Indicated by R Behind the Name of the Test.

Measure	$t_{\text{Neur-Nor}}$	$t_{\text{P-Nor}}$	$t_{\text{P-Neur}}$
Dark vision (5 mins)	1.94	**3.42	1.21
(10 mins)	1.20	*2.23	.88
(15 mins)	.99	1.77	.60
(20 mins)	1.34	1.85	.27
(30 mins)	**2.62	*2.53	.19
Total fusional reserves (Near Vision) R	1.06	*2.39	1.14
Total fusional reserves (Distant Vision) R	1.75	**2.89	.87
Accommodation (Binocular: blur)	1.87	**2.90	1.07
(Binocular: recovery)	*2.46	**3.63	1.33
(Right eye: blur)	*2.07	**2.95	.65
(Right eye: recovery)	*2.24	**3.63	.96
(Left eye: blur)	*2.24	**3.19	1.13
(Left eye: recovery)	*2.58	**4.04	1.83
Livingston Binocular gauge (Subjective convergence)	.16	.66	.45
Bishop Harman Diaphragm test R	.43	.78	.27
Autokinetic reaction time	1.00	1.66	1.26
Retinal rivalry R	**2.74	**5.08	1.19
Worth 4-dot reversals R	1.09	**2.79	.63
Street Gestalt completion test (correct responses) R	.32	1.69	1.02
Street Gestalt completion test (mean response time)	.95	1.57	.25
PMA space R	.65	**2.82	1.40
Peripheral span R	1.48	**4.47	1.34
Circles test (average response time)	.16	1.62	.92
Circles test (retardation in judgment time) R	.22	**2.69	.95
Kundt illusion (slow performance) R	1.55	*2.21	.77
Angular illusion R	1.72	*2.35	.34
Aubert phenomenon (total difference)	**2.61	1.57	.04
After-Image (duration of first image) R	1.78	**3.92	1.08
After-Image (average duration of two tests) R	1.56	**4.41	1.38
After-Image (1 - 2) R	1.45	1.46	.08
Prism glasses (adaptation score) R	.95	**3.09	*1.89
Lines assessment (orderly position)	.55	*2.31	*1.89
Lines assessment (random position)	1.49	**2.65	1.21
Learning dots (trials)	1.80	1.88	.11
Learning dots (wrong groups)	1.56	1.79	.15
Square pattern (error level) R	.02	.39	.31
Expressive movement (surface first drawing)	1.23	.09	.49
Expressive movement (average sidelength)	*2.00	**2.96	.44
Visual object recognition (head content)	1.68	*2.33	.51
Visual object recognition (spectacles content)	*2.33	*2.46	.06
Visual object recognition (head third dimension)	*2.76	**5.54	1.46
Visual object recognition (spectacles third dimension)	**2.60	**3.93	.21
Visual object recognition (head-spectacles content)	**2.22	**2.91	.42
Visual object recognition (head-spectacles third dimension: combined)	**3.55	**5.32	1.06
Lie scale R	.09	1.17	.75
Interests R	**3.00	**3.73	1.39
Dislikes R	1.82	**4.09	1.40
Sentence completion R	.97	1.51	.74
Mean log time on Nufferno test	1.12	**4.37	**3.38
Skew score	1.63	*2.46	1.82

*Significant at the 5% level.
**Significant at the 1% level.

Table 1 gives the results for those 50 tests on which the neurotics were intermediate between normals and psychotics. Table 2 gives the *t*-tests for 13 tests on

TABLE 2. TESTS WHERE THE MEAN OF THE NORMALS LIES BETWEEN THE MEAN OF THE NEUROTICS AND THE MEAN OF THE PSYCHOTICS. A REVERSED ORDER IS INDICATED BY R BEHIND THE SAME OF THE TEST.

Measure	$t_{\text{Neur-Nor}}$	$t_{\text{P-Nor}}$	$t_{\text{P-Neur}}$
Nufferno Intelligence Test R	1.15	.23	1.02
Giles Rule (Objective Convergence)	.60	1.71	1.88
Critical Flicker Frequency	*2.30	*2.25	**3.18
Gottschaldt figures (A) R	.59	*2.02	1.85
Gottschaldt figures (B) R	1.29	.77	1.63
Block design R	*2.48	.71	*2.16
Aubert phenomenon R	*2.29	.22	1.68
After-image (first distortion score) R	1.76	.51	1.73
Prism glasses R	1.36	.99	*1.96
Square pattern (trials) R	.53	.89	1.05
Word connection	.19	1.14	.94
Nufferno level R	1.15	.23	1.02
Error score	.65	1.30	1.51

*Significant at the 5% level.
**Significant at the 1% level.

which the normals were placed between the neurotics and the psychotics, while Table 3 lists the *t*-tests for 28 tests on which the mean for psychotics is between the

TABLE 3. THE MEAN OF THE PSYCHOTICS LIES BETWEEN THE MEAN OF THE NEUROTICS AND THE MEAN OF THE NORMALS. R INDICATES REVERSED ORDER.

Measure	$t_{\text{Neur-Nor}}$	$t_{\text{P-Nor}}$	$t_{\text{P-Neur}}$
Age	**2.85	2.84	.81
Dark vision (25 mins)	*2.04	*2.09	.11
Visual acuity (Snellen) R	**2.64	**3.06	.69
Farnsworth-Munsell test R	**4.26	1.72	1.24
Color Form attitudes (Ranking test) R	**2.82	**3.25	.54
K Test	1.74	1.64	.19
Necker cube	1.72	.94	.08
Dotted outlines R	.59	1.17	.56
Muller-Lyer illusion	**3.20	1.94	.91
Sander illusion	**3.29	1.42	1.34
Poggendorff illusion R	**3.25	*2.34	.56
Kundt illusion (fast performance) R	1.82	1.17	.62
Aubert phenomenon (Total - absolute volume)	1.61	.55	.57
After-image (fluctuation) R	**3.76	.51	**3.00
After-image (second distortion score)	.63	.16	.30
Emmert's Law (first screen stimulus)	1.25	1.22	.08
Emmert's Law (third screen stimulus)	.70	.60	.11
Prism glasses (average of four)	*2.36	.56	1.51
Eye-hand co-ordination	**2.73	*2.00	.42
Eye-hand co-ordination (second correction score)	**4.63	*2.32	1.05
Eye-hand co-ordination (adaptation score) R	1.48	.17	.79
Expressive movement (surface third drawing)	1.04	*2.32	.21
Expressive movement (surface difference)	**2.58	.61	.94
Expressive movement (sidelength difference)	**3.44	1.64	.52
Maudsley questionnaire	**5.21	**4.09	1.18
Food dislikes R	1.84	.70	.69
Annoyances	.24	.26	.02
Worries	**3.54	*2.25	1.40

*Significant at 5% level.
**Significant at 1% level.

mean for the normals and the mean for the neurotics. Table 4 summarizes the results. Column A shows the number of tests which discriminated at the 5 per cent level of significance and column B shows the number of tests which discriminated at the 1 per cent level of significance.

TABLE 4. SUMMARY OF THE TESTS DISCRIMINATING BETWEEN NORMAL, NEUROTIC AND PSYCHOTIC GROUPS AT 2 LEVELS OF SIGNIFICANCE

Groups Compared	Groups With Intermediate Mean	A Number of Tests $p \leqq .05$	B Significant $p \leqq .01$	Number n. s.
Psychotics - Normals	Neurotics	9	25	16
	Normals	2	0	11
	Psychotics	6	3	19
Neurotics - Normals	Neurotics	8	7	35
	Normals	3	0	10
	Psychotics	2	14	12
Psychotics - Neurotics	Neurotics	2	1	47
	Normals	2	1	10
	Psychotics	0	1	27

The findings that on 41 out of 81 tests the mean for the neurotics did not occupy a position intermediate between the normals and psychotics gives experimental support to Eysenck's theory that neuroticism and psychoticism form two independent continua. Unfortunately, the data indicate that out of a list of some 91 tests scores only 7 discriminated at the 5 percent level of significance between psychotics and neurotics, and of these 7 tests only one test (CFF) discriminated all 3 types of subjects from one another.

SUMMARY

Data presented by Eysenck *et al*[1] have been reanalyzed to obtain *t*-tests of the statistical significance of the differences between the three groups of subjects. When a 5 per cent level of significance is used, we find that 44 tests discriminated between psychotics and normals and 34 tests discriminated between neurotics and normals, but only 7 tests discriminated between psychotics and neurotics. CFF was the only test which discriminated all three types of subjects. The results tend to support Eysenck's bi-factor theory of neurosis *vs.* psychosis, but show that many of the tests lack discriminatory power.

REFERENCES

1. EYSENCK, H. J., GRANGER, G. W. and BRENGELMANN, J. C. Perceptual Processes and Mental Illness. *Maudsley Monogr. Series # 2*, Chapman and Hale, London, 1957.

HOMOSEXUALITY AND CREATIVITY

ALBERT ELLIS

333 West 56th Street
New York, N. Y.

PROBLEM

Although there is much speculation about homosexuals being more artistic or creative than heterosexuals, a search of the literature discloses no objective studies in this connection. Indeed, studies of homosexuals which employ an adequate control group of heterosexuals, and which compare any of their characteristics or aspects of their behavior with those of a control group, are very rare. This study compares a group of homosexual patients in psychotherapy with an equated group of highly heterosexual psychotherapy patients to determine group differences in clinical diagnosis, creativity, and clinical improvement during therapy.

Method

Sixty-six individuals with severe homosexual problems and 150 exceptionally heterosexual psychotherapy patients, who were seen by the writer for a minimum of ten and a maximum of 350 sessions, were equated for age, sex, and educational status. The homosexual patients were further divided into three subgroups: (a) 9 male and 10 female bisexuals (that is, individuals with a strong homosexual component, including a history of overt homosexuality, but who also had a reasonably strong heterosexual component); (b) 23 male and 10 female fixed homosexuals (that is, individuals who were exclusively or predominantly homosexual but who maintained their own sex role); (c) 10 male and 4 female inverts (that is, individuals who were exclusively homosexual and who maintained what Brown[1] has called sex role inversion, with the males adopting a female or "fairy" role and the females playing a masculine or "butch" role in their everyday living.

Each of the patients was rated by the therapist, on the basis of his intimate psychotherapeutic experience with the subject, as to whether he or she was (a) neurotic, borderline psychotic, or psychotic; (b) highly creative, moderately creative, or non-creative; (c) distinctly more creative or little more creative following therapy; (d) considerably, distinctly, or little improved clinically during therapy; and (e) considerably, distinctly, or little changed in the direction of heterosexual behavior during therapy[4]. In the ratings for creativity, the criteria hypothesized by Guilford[6] ("sensitivity to problems, ideational fluency, flexibility of set, ideational novelty, synthesizing ability, analyzing ability") and English and English[5] ("to find new solutions to a problem or new modes of artistic expression") were employed.

Results

Table 1. Comparisons Between Exceptionally Heterosexual and Homosexual Psychotherapy Patients

Variable	Exceptionally Heterosexual Patients		Homosexual Patients: Bisexuals		Fixed Homosexuals		Homosex. Inverts		Total Homosexual	
	N	%	N	%	N	%	N	%	N	%
Clinical diagnosis										
Neurotic	117	78%	12	63%	20	59%	6	43%	38	59%
Borderline psychotic	27	18%	4	21%	10	32%	5	36%	19	28%
Psychotic	6	4%	3	16%	3	9%	3	21%	9	13%
Total	150	100%	19	100%	33	100%	14	100%	66	100%
Creativity										
Highly creative	39	26%	2	11%	3	9%	0	0%	5	8%
Moderately creative	47	31%	10	52%	13	39%	2	14%	25	38%
Non-creative	64	43%	7	37%	17	52%	12	86%	36	54%
Increase in creativity following therapy										
Distinct increase	92	61%	12	63%	18	54%	4	28%	34	52%
Little or no increase	58	39%	7	37%	15	46%	10	72%	32	48%
Clinical improvement during therapy										
Considerable	77	51%	9	47%	8	22%	1	7%	18	27%
Distinct	60	40%	7	37%	15	46%	4	29%	26	39%
Little or none	13	9%	3	16%	10	32%	9	64%	22	34%
Clinical improvement in relation to homosexual problem during therapy										
Considerable			12	63%	9	27%	1	7%	22	34%
Distinct			7	37%	9	27%	3	21%	19	28%
Little or none			0	0%	15	46%	10	72%	25	38%

The results of the therapist's ratings on the 66 homosexual and 150 highly heterosexual patients used in this study are shown in Table 1. When some of the observed findings in this table are subjected to Chi-square analysis the following significant and non-significant differences are obtained:

1. The exceptionally heterosexual subjects are significantly more often diagnosed as neurotic and less often diagnosed as borderline psychotic and psychotic than the exceptionally homosexual (that is, fixed homosexual and inverted) subjects ($p = .01$).

2. The exceptionally heterosexual subjects are significantly more often rated as highly or moderately creative and less often rated as non-creative than the exceptionally homosexual subjects ($p = .01$).

3. With clinical diagnosis held constant, the neurotic heterosexual subjects are more often rated as highly or moderately creative and less often rated as noncreative than the fixed homosexual and inverted subjects; but this difference is not statistically significant.

4. The extremely heterosexual subjects are more often rated as showing a distinct increase in creativity following psychotherapy than the extremely homosexual subjects; but this difference is not statistically significant.

5. The exceptionally heterosexual subjects are significantly more often rated as showing considerable or distinct clinical improvement and less often rated as showing little or no clinical improvement during therapy than the exceptionally homosexual subjects ($p = .01$).

6. With clinical diagnosis held constant, the neurotic heterosexual subjects are significantly more often rated as showing considerable or distinct clinical improvement than the neurotic fixed homosexuals and inverts ($p = .01$).

7. Compared to the fixed homosexual and inverted subjects, the bisexual subjects are more frequently diagnosed neurotic and less frequently borderline psychotic and psychotic (non-significant); significantly more often rated highly or moderately creative and less often rated non-creative ($p = .06$); more often rated as having increased their creativity following therapy (non-significant); significantly more often rated as showing considerable or distinct clinical improvement during therapy ($p = .05$); and significantly more often rated as showing considerable or distinct gains in heterosexuality during therapy ($p = .01$).

8. Compared to the homosexual inverts, the fixed homosexuals without sex role inversion are more often diagnosed as being neurotic and less often as being borderline psychotic or psychotic (non-significant); significantly more often rated as being highly or moderately creative ($p = .05$); more often rated as having increased their creativity following therapy (non-significant); significantly more often rated as showing considerable or distinct clinical improvement during therapy ($p = .05$); and more often rated as showing considerable or distinct gains in heterosexuality during therapy (non-significant).

Discussion

The obtained data very consistently, though at times non-significantly, uphold the hypothesis that degree of homosexuality is directly related to both degree of emotional disturbance and loss of creative potential. This hypothesis would also tend to be supported by the findings of Drevdahl[2], who found that although creative individuals tend to be more non-conforming than non-creative ones, they also tend to be more self-sufficient and flexible. In a follow-up study, Drevdahl and Cattell[3] found that creative artists and writers differed from the normal population in being more emotionally mature, dominant, adventurous, and less cyclothymic, surgent, and subject to group standards and control. The findings of the present and these confirmatory studies would tend to question seriously the long held stereotype of the creative individual as neurotic or psychotic.

It may well be, as Drevdahl's original study indicates, that if we limit the term "creative" to individuals working in artistic or esthetic fields, such individuals are somewhat more suspicious and disturbed than are other creative individuals (such as scientists) or non-creative persons. But if we include under the rubric of creativity all kinds of inventive, original individuals, the traditionally held stereotype quickly breaks down.

In the case of homosexuals in particular, it may well be that the illusion that is commonly held of their being unusually creative individuals results from such factors as (a) their frequently gravitating to esthetic rather than scientific or industrial fields; (b) the ease of one's gaining popular fame or notoriety, in our culture, when one is prominent in esthetic or artistic rather than other creative areas; (c) the common confusion which frequently exists between an artistic *performance* (as in the theatre, the opera, or the dance) and an original *creation* (e.g., a fine play, invention, or industrial technique); and (d) the fact that not a few homosexual creators become unusually well-known because of their aberrated lives, while many heterosexual originators live relatively unobtrusive and relatively unknown lives.

Summary

Sixty-six individuals with severe homosexual problems and 150 exceptionally heterosexual psychotherapy patients were rated by the therapist as to clinical diagnosis, creativity, increase in creativity following therapy, clinical improvement, and (in the case of the homosexuals) increase in heterosexuality. The main findings were:

1. Compared to the homosexual patients, the exceptionally heterosexual subjects were significantly less emotionally disturbed, more creative, and more clinically improved during therapy.

2. Compared to the fixed homosexual and inverted patients, the bisexual subjects were significantly more creative, more clinically improved during therapy, and made greater gains in heterosexuality.

3. Compared to the homosexual inverts studied, fixed homosexuals without sex role inversion were significantly more creative and more clinically improved during therapy.

4. It appeared likely that a good part, though probably not all, of the obtained relationships between degree of homosexuality, on the one hand, and lesser creativity and clinical improvement with therapy, on the other hand, was attributable to the greater severity of emotional disturbance found among the homosexual patients.

References

1. Brown, D. G. Transvestism and sex role inversion. In A. Ellis and A. Abarbanel, (Eds.), *Encyclopedia of sexual behavior*. New York: Hawthorn Books, 1959.
2. Drevdahl, J. E. Factors of importance for creativity, *J. clin. Psychol.*, 1956, *12*, 21-26.
3. Drevdahl, J. E., and Cattell, R. B. Personality and creativity in artists and writers. *J. clin. Psychol.*, 1958, *14*, 107-111.
4. Ellis, A. The effectiveness of psychotherapy with individuals who have severe homosexual problems. *J. consult. Psychol.*, 1956, *20*, 191-195.
5. English, H. B., and English, Ava C., *A comprehensive dictionary of psychological and psychoanalytical terms*. New York: Longmans, Green, 1958.
6. Guilford, J. P. Creativity. *Amer. Psychol.*, 1950, *5*, 444-454.

JUDGING THE RELATIVE SEVERITY OF PSYCHIATRIC OUT-PATIENT COMPLAINTS[1]

K. WARNER SCHAIE
University of Nebraska
LOIS A. ROMMEL AND JAMES M. A. WEISS
Washington University School of Medicine and Malcolm Bliss Research Laboratories

Problem

Disposition of a patient entering a psychiatric out-patient clinic depends to a large extent upon the evaluation made of the patient's complaints during the initial screening procedure. One important consideration in such disposition will undoubtedly be the nature and relative severity of the presenting complaints. The purpose of the present study is to examine whether there are stimulus dimensions among the multitude of possible psychiatric complaints which lead to differential judgment of the severity of such complaints on the part of the intake interviewer. An attempt will also be made to see whether a classification of psychiatric complaints can be made in terms of a scheme that is meaningful to professional raters from three mental health professions and to test the hypothesis that judgmental differences may be a function of difference in professional affiliation. While it is known that there are differences of opinion as to the characteristics of personality disturbance among members of these professional groups (1), it remains to be seen whether such differences also will be reflected in the process of assessing the severity of patient complaints. To overcome the difficulty of adequately sampling widely diverse professional groups, use is made of the replication of small samples and of obverse factor analysis.

Procedure

In the course of another study approximately 200 successive case files from a psychiatric out-patient clinic were examined and all presenting complaints noted (6). These complaints were then sorted into different categories until a meaningful system of classification emerged (10). In contrast to studies such as that by Lorr and Rubinstein (4), or Tatom (8), the experimenters were not concerned with classification into a given nosological system of psychiatric diagnosis, but rather with the meaningful assignment of verbalized complaints to a system on which agreement by professional raters would be possible. Two dimensions were found to be required to place each complaint. These dimensions were named the "determinants" and "referents" of a given complaint. The determinant identifies the specific nature of the complaint while the referent gives the attributed cause or result to which the nature of the complaint refers. Seven determinants and five referents were found relevant for classification and were named as follows:

Determinants	Referents
1. Affective	a. Physical health
2. Anxietal	b. Mental health
3. Behavioral	c. Economic
4. Mentational	d. Inter-personal
5. Reality distortive	e. Non-specific
6. Welfare	
7. Somatic	

The combinations of determinants and referents provide 35 possible classification units. An example for a complaint which fits category 1a would be: "I feel blue because my health is bad." The determinant of this complaint is of an "affective"

[1]This paper is one of a series reporting a larger study of mental health problems of older persons in progress at Malcolm Bliss Psychiatric Clinic, St. Louis, Missouri. It is supported by Research Grant M-1252 from the National Institute of Mental Health, U. S. Public Health Service.

nature (*i.e.* I feel blue) and names as its cause problems referring to "physical health".[2] All complaints were then sorted into these categories and a balanced Q-sort sample was composed by selecting two representative complaints from each category. This gave a total of seventy complaints. Professional raters were asked to distribute the seventy complaints in a quasi-normal Q-sort distribution containing 11 intervals.[3] This method is a convenient procedure for ranking a relatively large number of items along a continuum using manageable numbers of discriminative choices [7].

Raters were instructed to distribute the complaint items in their order of severity. They were asked to place the two items which they considered to be the most disabling complaints into the category identified by a score of ten and those two items which were least disabling into the category which yielded a score of 0. This procedure continued until all items had been assigned according to the required distribution.

The use of Q-methodology permits tests of certain types of hypotheses even though only few raters are used. The basic assumption made is that one is dealing with a population of complaints rather than raters. While there may be considerable sampling differences among raters with respect to any given complaint, one may still propose with some confidence that these differences will be in terms of dimensions which may be found in professional raters in general or which are confined to raters belonging to a specific professional group. Instead of collecting a large number of ratings with respect to given complaints and then testing the hypothesis concerning the presence or absence of significant mean differences between two professional groups, one may instead collect statements (or complaints) about hypothesized dimensions and ask individual raters to rank these propositions. The rankings given by the raters may then be correlated and factored to see whether there is a unique dimension in the judgment of complaint characteristics which is used by members of a given professional group but not by others. The analysis of variance may be used to test the hypothesis that such distinct dimensions do in fact exist and whether they are reflected by significant differences among the mean scores yielded by our system of classification.

Two sets of raters, each consisting of three psychiatrists, three psychologists and three social workers, were utilized. The sets were independently drawn from available professional persons in two geographically separated areas to permit assessment of the stability of results based on small sample statistics by means of replicated samples. All the participant raters had the standard qualifications of their respective professions: psychiatrists, M. D. plus eligibility for Board certification in psychiatry; psychologists, Ph.D.; social workers, M. S. W. All were engaged in some kind of activity which required dealing with psychiatric patients, although the various raters were affiliated with different institutions.

Results

Hypothesis-testing factor analysis. As a first step in the analysis, the intercorrelations among raters were computed and are presented in Table 1, showing significant agreement among most of the raters. The greatest intra-group agreement was found among the psychologists. It is of interest, also, that several of the psychiatrists and social workers showed greater agreement with psychologists than with other members of their own professional group.

The next step was to factor the inter-correlation matrices for the two separate sets of ratings. A hypothesis-testing factor analysis procedure, with replication to provide a basis for generalization, was employed for this purpose.[4] If there is some

[2]Details of the procedure for selecting the final complaint categories are reported in [10].

[3]The number of items which were required to be sorted into each interval are as follows:

Score value of interval	0	1	2	3	4	5	6	7	8	9	10
No. of items placed in interval	2	3	5	8	11	12	11	8	5	3	2

[4]In this application of factor analysis, the experimenter groups those variables which are hypothesized to have significant loadings on a specific factor and then proceeds to test whether the proposed structure can be found. The method used is a modification of Thurstone's multiple group method [9] and is more fully described in a paper by Horst and Schaie [3].

TABLE 1. INTERCORRELATIONS BETWEEN RATERS (N = 70 COMPLAINT STATEMENTS)*

		Psychi 1	Psychi 2	Psychi 3	Psycho 1	Psycho 2	Psycho 3	S.W. 1	S.W. 2	S.W. 3
		Set A								
	Psychiatrist 1		.05	.39	.24	.30	.29	.27	.22	.24
	" 2			.07	.17	.17	.08	.28	.22	−.09
	" 3				.61	.74	.66	.61	.44	.49
Set A	Psychologist 1					.66	.51	.57	.50	.48
	" 2						.60	.71	.53	.70
	" 3							.50	.33	.45
	Social Worker 1								.45	.43
	" 2									.51
	" 3									
		Set B								
	Psychiatrist 1	−.14	.25	.28	.17	.34	.16	.14	.16	−.06
	" 2	.34	.26	.22	.31	.32	.34	.38	.28	.21
	" 3	.07	.52	.41	.61	.64	.51	.41	.52	.20
Set A	Psychologist 1	.14	.53	.52	.57	.66	.48	.48	.54	.30
	" 2	.20	.62	.52	.65	.71	.55	.44	.56	.27
	" 3	.10	.47	.32	.51	.60	.48	.39	.41	.21
	Social Worker 1	.06	.62	.53	.63	.69	.55	.50	.64	.41
	" 2	.34	.61	.49	.47	.47	.53	.47	.42	.19
	" 3	.26	.60	.52	.46	.50	.44	.32	.41	.18
		Set B								
	Psychiatrist 1		.50	.44	.38	.24	.48	.38	.23	.38
	" 2			.69	.66	.62	.71	.59	.68	.39
	" 3				.70	.59	.59	.55	.63	.37
Set B	Psychologist 1					.66	.69	.69	.75	.33
	" 2						.54	.63	.68	.30
	" 3							.50	.64	.45
	Social Worker 1								.70	.34
	" 2									.45
	" 3									

*Correlations are significant at the 1% level of confidence if they exceed .30, and at the 5% level if they exceed .23.

variance which is specifically attributable to professional identification, regardless of any other factors operating in the judgment of the seventy psychiatric complaints, then by maximizing the variance common to the members of each professional triplet and by minimizing that portion of the variance common to both the members of the triplet and all the other members of the set of raters, one should be able to obtain a solution where in a three factor system (one for each profession) simple structure can be shown to exist. That is, if, and only if, there is a unique factor contributing to the manner in which members of the different professions rank psychiatric complaints, then the factorial solution will contain one factor for each group on which the members of the professional triplet have positive and significant loadings, while all other raters have zero or chance loading.

Table 2 presents the factor loading matrices before and after oblique rotation. If one specifies that a factor loading below .20 is to be considered a chance deviation from zero, then examination of this table leads to the conclusion that the hypothesis as to the presence of an unique judgmental factor related to the professional groupings is not adequately supported. There is overlap on each of the three factors in at least one of the two sets and it is clear that this solution is not the most parsimonious one. On both sets, however, there seems to be some unique factor saturation for some of the psychiatrists and the social workers, while the hypothesized "psychologist factor ' would seem to represent a dimension of inter-professional communality.

TABLE 2. FACTOR-LOADINGS FOR THE HYPOTHESIS-TESTING FACTOR ANALYSIS*

	Multiple Group Factor Loadings			Factor Loadings After Oblique Rotation		
	Set A					
	I	II	III	I	II	III
Psychiatrist 1	.49	.35	.03	.37*	−.24*	.22*
Psychiatrist 2	.23	.18	.06	.14	−.10	.16
Psychiatrist 3	.01	.85	−.21	.15	.38*	.00
Psychologist 1	−.06	.77	−.01	−.03	.44*	.16
Psychologist 2	.11	.85	.29	−.08	.37*	.50*
Psychologist 3	−.05	.75	−.28	.15	.37*	−.10
Social Worker 1	.19	.75	.09	.10	.23*	.31*
Social Worker 2	.19	.58	.41	−.10	.19	.57*
Social Worker 3	.48	.69	.31	.19	−.02	.57*
	Set B					
Psychiatrist 1	.52	.45	.06	.49*	−.05	.01
Psychiatrist 2	.23	.82	−.02	.35*	.30*	−.01
Psychiatrist 3	.24	.77	−.04	.35*	.29*	−.03
Psychologist 1	−.02	.86	.01	.12	.40*	.06
Psychologist 2	−.13	.77	.04	.00	.37*	.10
Psychologist 3	.15	.80	−.05	.29*	.35*	−.03
Social Worker 1	−.01	.75	.25	.00	.14	.29*
Social Worker 2	−.20	.85	.20	−.12	.31*	.27*
Social Worker 3	.24	.44	.39	.08	−.20*	.37*

*Factor loadings greater than .20 are considered to be significant.

Analysis of variance and examination of mean scores. The analysis of variance was next used to test hypotheses with respect to the meaningfulness of our classification system and to assess further the effect (if any) of membership in a specific profession upon the rankings made. Since the scores on the different categories were all obtained by Q-sort procedure, the variance for individual raters and professional groupings are, of course, required to be equal by definition. An overall analysis of variance test is possible, however, although group and individual differences must be tested by examining the simple interactions.[5]

The applicable null hypotheses for this analysis may be stated as follows: (a) there are no differences between the various complaint determinants; (b) there are no differences between the various complaint referents; (c) there is no significant interaction between complaint determinants and referents; (d) there are greater differences among raters *within* each professional group than among the groups on the complaint determinants; (e) there are greater differences among raters *within* each professional group than among groups on the complaint referents; (f) the interaction between determinants and referents *within* each professional group is larger than the interaction between the complaint dimensions among the groups.

Table 3 shows the summary for the analysis of variance and indicates that hypotheses a, b and c must be rejected. Significant overall differences among the complaint determinants, referents and their interaction suggests that the classification system is reasonable and useful in distinguishing between different types of com-

[5]The analysis of variance procedure used is based upon a mixed model, where determinants, referents and professions are assumed to be fixed constants while the raters are considered a random variate[(3)].

TABLE 3. ANALYSIS OF VARIANCE TESTING SIGNIFICANCE OF EFFECTS OF VARIATION AMONG PROFESSIONAL GROUPS AND RATERS OVER ALL REFERENTS AND DETERMINANTS

Source of Variation	Sum of squares				Degrees of Freedom*	Mean squares		F-ratios	
	Set I		Set II			Set I	Set II	Set I	Set II
Referents		205.79		174.32	4	51.45	43.58	12.05***	8.56***
Raters + Referents		136.78		162.82	32	4.27	5.09		
Professions + Referents	49.54		37.25		8	6.19	4.66	1.70	.89
Pooled Raters + Referents within professions	87.24		125.57		24	3.64	5.23		
Determinants		960.20		723.13	6	160.03	120.52	25.28***	18.86***
Raters + Determinants		304.00		306.87	48	6.33	6.39		
Professions + Determinants	74.87		53.80		12	6.24	4.48	.98	.64
Pooled raters + Determinants within professions	229.13		253.07		36	6.36	7.03		
Referents + Determinants		227.01		545.11	24	9.46	22.71	2.48***	9.31***
Raters + Refer. + Determ.		732.72		468.75	192	3.82	2.44		
Profess. + Ref. + Determ.	187.26		163.06		48	2.68	3.40	.71	1.60
Pooled raters + Ref. + Determ. within Professions	545.46		305.69		144	3.79	2.12		
Within Cells		781.50		967.00	315	2.48	3.07		
Total Variance		3348.00		3348.00	621				

*Each set has total degrees of freedom: 629 - 8 (8 df. lost because of lack of differences among raters due to the forced distribution used in the Q-sorts). **Significant at 1% level. *** at .001 level. Trivial F-ratios are omitted.

plaints. Hypotheses d, e and f are accepted and suggest the conclusion that there are no significant differences among the professional groupings in assessing the severity of either specific complaints, complaint determinants or complaint referents. The results for this analysis are shown to be essentially replicated for the second set of raters. In both instances, the hypotheses with respect to the meaningfulness of the classification system are confirmed and the hypotheses with respect to interprofessional differences rejected.

What then are the dimensions underlying the judgment of a complaint's severity? A clue may be found in the mean scores for each of the *a priori* sorting classifications. Tables 4 and 5 give the mean scores for each of the complaint determinants

TABLE 4. MEAN SCORES OF RATERS ON THE COMPLAINT REFERENTS
(N = 14 STATEMENTS FOR EACH OF 6 RATERS IN EACH GROUP)

	Psychi.	Psychol.	Soc. Work.	All Raters
Physical health	5.02	4.72	4.63	4.79
Mental health	5.37	5.67	5.73	5.59
Economical	4.21	3.74	4.27	4.08
Interpersonal	5.27	5.30	5.54	5.37
Nonspecific	5.12	5.57	4.83	5.18

and complaint referents. Since the initial analysis suggests that the two sets of raters are quite similar, their ratings were now pooled and mean scores are given for the six members of each professional grouping as well as for all eighteen raters. Using the within cells variance over all the raters (from the analysis of variance) the stand-

TABLE 5. MEAN SCORES OF RATERS ON THE COMPLAINT DETERMINANTS
(N = 10 STATEMENTS FOR EACH OF 6 RATERS IN EACH GROUP)

	Psychi.	Psychol.	Soc. Work.	All Raters
Affective	4.80	4.90	5.07	4.92
Anxietal	4.40	4.55	4.58	4.72
Behavioral	4.90	4.40	4.72	4.67
Mentational	5.33	5.97	4.75	5.35
Reality distortive	7.02	7.77	7.63	7.47
Welfare	3.97	3.17	3.65	3.59
Somatic	4.58	4.25	4.60	4.48

ard error of the difference for scores on the complaint determinants was found to be .30 and for the complaint referents to be .26. Using the 3 S. D. criterion, significant differences must attain a magnitude of .90 score points for determinants and .78 score points for the referents. Only one determinant shows a significant difference, the psychologists rating "mentational" complaints more severe than did the social workers. None of the referents showed significant differences.

When one examines the mean differences among categories for all the raters, however, more significant findings occur. The standard errors of the difference for this analysis will be .18 and .15 and the significant score magnitudes (at the 3 S. D. level), .54 and .45 for the complaint determinants and referents, respectively. All raters agree that "reality distortive" complaints are to be considered most severe. "Mentational" complaints are next in severity and receive a significantly higher mean score than all other determinants except "reality distortive" and "affective". "Welfare" complaints (*i.e.*, asking for aid of a non-psychiatric type) are rated significantly less important than all other determinants. Among the referents "mental health" and "interpersonal" receive the highest mean scores, both of which are significantly higher than "physical health". "Economic" is rated significantly less severe than all the other complaint referents.

A simple structure solution. Since there are considerable differences among individual raters in assessing the relative severity of each class of complaints it was hypothesized that there might be other subsidiary dimensions which would affect such ratings. To get at these subsidiary dimensions the inter-correlations among all raters (see Table 1) were factored by the centroid method. Five factors were extracted and rotated by the radial method with a view towards achieving simple structure within a positive manifold [2]. Table 6 lists both the centroid and the oblique factor loadings.

TABLE 6. CENTROID ANALYSIS OF THE INTER-CORRELATION AMONG ALL RATERS

	Centroid Factor Loadings					Oblique Factor Loadings				
	A	B	C	D	E	A	B	C	D	E
Psychiatrist 1	.31	.36	−.11	−.26	.10	−.19	.47	.06	−.02	−.11
Psychiatrist 2	.33	−.32	.15	−.15	−.16	.33	−.15	.21	.42	.27
Psychiatrist 3	.72	.42	.10	.06	−.10	−.21	.65	.35	.03	.14
Psychiatrist 4	.41	−.53	−.26	.14	−.16	.68	−.14	−.11	.38	.31
Psychiatrist 5	.84	−.11	−.22	.06	.02	.41	.39	.16	.22	.13
Psychiatrist 6	.76	−.19	−.21	−.15	.19	.45	.29	.19	.22	−.04
Psychologist 1	.72	.23	.12	−.09	−.04	−.04	.50	.38	.14	.11
Psychologist 2	.81	.35	.00	.10	−.02	−.08	.67	.32	.00	.08
Psychologist 3	.63	.34	.14	.20	−.18	−.18	.54	.32	.02	.21
Psychologist 4	.84	−.14	.12	−.03	−.06	.32	.25	.42	.33	.21
Psychologist 5	.83	.12	.18	−.11	.02	.06	.44	.50	.19	.10
Psychologist 6	.78	−.22	−.08	.18	−.11	.44	.23	.21	.29	.27
Social Worker 1	.76	.13	.25	.02	.25	.01	.39	.60	−.05	−.14
Social Worker 2	.65	.05	−.30	−.10	−.18	.24	.46	−.05	.32	.27
Social Worker 3	.63	.28	−.38	.17	.15	.06	.66	−.02	−.16	−.10
Social Worker 4	.72	−.25	.18	.26	−.07	.37	.08	.42	.22	.23
Social Worker 5	.79	−.20	.22	−.08	.20	.32	.15	.58	.19	−.04
Social Worker 6	.45	−.32	.11	.30	.13	.37	−.08	.32	.02	.00

Factor E seems to collapse upon Factor A, and an attempt at interpretation is therefore made for the first four factors only.

Factor arrays were computed for the four factors by selecting the three ratings which had the highest oblique loading on each factor. All statements were then reordered in terms of these factor arrays. Table 7 lists the five most extreme statements in each factor array. All the arrays contain statements which have "reality distortive" aspects. Thus, as already noted, the common aspect of ranking complaint

TABLE 7. MOST EXTREME STATEMENTS IN TERMS OF THE FACTOR ARRAYS

Factor A
1. I have unusual ideas about what is happening to my body.
2. Someone is trying to harm me.
3. I am so nervous I can't work.
4. I want something done because I am in poor health.
5. I want to go to a hospital because I am upset.

Factor B
1. I believe I have special powers.
2. I hear or see things no one else does.
3. I get strange ideas when I get nervous.
4. I have unusual ideas about what is happening to my body.
5. Strange things happen to me when I get upset.

Factor C
1. I hear or see things no one else does.
2. Someone is trying to harm me.
3. I believe I have special powers.
4. I feel I'm a burden to my family. Maybe they would be better off without me.
5. I worry about losing my mind.

Factor D
1. I am so nervous I can't work.
2. I lose jobs because of the way I act.
3. I have unusual ideas about what is happening to my body.
4. I have trouble at work because my mind wanders and I get confused.
5. Someone is trying to harm me.

statements is in terms of a dimension reaching from reality to irreality, the latter being considered the more severe. In terms of the factor arrays we can identify four additional dimensions. Factor A seems to reflect that "desire of the patient to be taken care of" is the basis of judging severity of complaints. Factor B seems to be concerned with the "distortion of the self-concept"; Factor C with the "disruption of interpersonal relations;' and Factor D with the "economic problem" associated with the respective complaints.

In the light of the earlier analysis of possible professional differences, it is of interest to note that four of the six psychiatrists have significant loadings on Factors A and B, while four of the six psychologists have significant loadings on Factor B and five of the six on Factor C. This finding might suggest more concern of the psychiatrists with the patient's desire to receive help as a significant variable in a complaint's severity and more concern of the psychologists with the disruption of interpersonal ties. Members of all three professions, however, have loadings on all the factors, underlining the tentative conclusion that individual value systems rather than professional affiliation are of major importance in mediating the judgment of the severity of a psychiatric out-patient complaint.

Summary and Conclusions

A balanced sample of seventy complaint statements typical of patients visiting a psychiatric out-patient clinic was selected from a larger sample of complaints culled from clinic files. Two sets of raters, each consisting of three psychiatrists, three psychologists and three social workers rank-ordered these complaints in the order of their severity using Q-sort methods.

All complaints were classified in terms of a two-dimensional system of determinants and referents. The results of an analysis of variance indicates that there were significant differences among individual raters, but that there was more overall divergence *within* each professional group than *among* individuals from different professional backgrounds. Complaints referring to mental health, interpersonal, or nonspecific sources were felt to be most significant by all raters, while those referring to economic and physical health dimensions were rated least severe. In terms of the determinants, complaints of a reality distortive or of mentational nature were held to be most severe, while those of a welfare or somatic nature were considered least severe.

The rankings made by the members of each set of raters were correlated and factored by the multiple-group method to test the hypothesis that there are unique factors associated with the professional identification of the raters which will influence their judgment of the severity of psychiatric complaints. No such unique factor could be identified.

In a second factor analysis of the correlations among all the raters, using the centroid method, an oblique rotation resulted in four identifiable factors. These were named "desire of the patient to be taken care of', "distortion of the self-concept", "disruption of inter-personal relations" and "economic problems". Again, members of all professional groups show loadings on all factors.

If there are indeed unique stimulus dimensions operating in the judgment of the severity of psychiatric complaints which are related to professional identification, then these dimensions should emerge in a factor analysis design such as the one used here for any set of raters. The characteristics of a particular rater would effect primarily the magnitude of the specific factor loadings but not the simple structure solution if such unique dimensions exist. The results suggest that at least for

samples of complaint statements gathered in a municipal out-patient clinic a solution specifying effect of professional identification on the judgment of the severity of psychiatric complaints is not parsimonious and that there was more communality than divergence among the raters participating in the study. Instead, evidence is reported which suggests that such judgment may be mediated by the raters' individual value systems with respect to the nature of the specific complaints.

References

1. Glosser, H. T. Psychiatric versus psychological opinion regarding personality disturbances. *Amer. Psychol.*, 1958, *13*, 477-481.
2. Guilford, T. P. *Psychometric methods.* New York: McGraw-Hill, 1954.
3. Horst, P. and Schaie, K. W. The multiple group method of factor analysis and rotation to a simple structure hypothesis. *J. exper. Educ.*, 1956, *24*, 231-237.
4. Lorr, M. and Rubinstein, E. A. Factors descriptive of psychiatric out-patients. *J. abnorm. soc. Psychol.*, 1955, *51*, 514-522.
5. McNemar, Q. *Psychological Statistics.* New York: Wiley, 1956.
6. Rommel, Lois A., Schaie, K. W., Weiss, J. M. A. and Melikan, T. P. The relationship between medical, psychological and social factors and the psychiatric complaints of older patients. *Amer. Psychol.*, 1958, *13*, 341-342.
7. Stephenson, W. *The study of behavior: Q-technique and its methodology.* Chicago: University of Chicago Press, 1953.
8. Tatom, Mary H. A factorial isolation of psychiatric out-patient syndromes. *J. consult. Psychol.*, 1958, *22*, 73-81.
9. Thurstone, L. A multiple group method of factoring the correlation matrix. *Psychometrika*, 1945, *10*, 73-78.
10. Weiss, J. M. A., Rommel, Lois A. and Schaie, K. W. Presenting problems of older patients referred to a psychiatric clinic. *J. Gerontology* (in press).

THE COMPARATIVE RELIABILITY OF TWO TECHNIQUES OF PERSONALITY APPRAISAL[1]

PAUL HORST AND CALVIN E. WRIGHT

University of Washington

One of the traditional objections to the self-appraisal type of noncognitive psychological instrument is the tendency for the subject to give what he regards as the socially approved response to the items or statements on the inventory. A method credited by Baier(1) to the senior author has been used extensively in an effort to overcome this objection. It consists essentially of the following steps:

1. Several or more dimensions are postulated for a given personality, temperament, interest, or other noncognitive domain.

2. Sets of items assumed to be relevant to the respective postulated dimensions are written.

3. The extent to which agreement with each item is regarded as a socially acceptable response is determined by some scaling technique considered appropriate by the investigator. For each item, then, a scale value of social desirability is available.

4. Pairs of items are next arranged so that the two items have approximately the same social desirability scale value but represent different dimensions. Typically each item is paired with the same number of items as is every other item, so that on the final format all items appear the same number of times.

5. A subject is then required for each pair of items to indicate which of the two most nearly applies to him. The score for each dimension or trait is the number of items from that trait checked by the subject.

---

[1]The results of this study were a portion of the junior author's doctoral dissertation (1957) under the supervision of the senior author. The larger study was supported in part by Office of Naval Research Contract Nonr-477(08) and Public Health Research Grant M-743(C2), Paul Horst, Principal Investigator. A technical report form of the complete study has been distributed. Reproduction in whole or in part is permitted for any purpose of the United States Government.

This procedure has come to be known as the forced choice technique. Well known variations of it consist of arranging items in triads or tetrads, rather than in duads or pairs. The senior author's original notion in proposing the technique was that, since the items were matched for social desirability, the subject would be forced to choose on the basis of his relative position with respect to the two dimensions represented by the pair. He would therefore be responding to the dimensionality or trait aspects of the items rather than to what he regarded as socially desirable.

Another advantage of the forced choice technique is that a relatively small number of items for each dimension can yield a wide range of scores for that dimension. For example, if there are nine items for a dimension and each is paired with three other items, the range of possible scores for a dimension or trait is from 0 through 27. It should, however, be noted that this feature of the technique, while economical in item writing time, is not economical in test administration time, since the subject must respond to each member of all of the pairs in the format.

A serious defect in the forced choice technique was early recognized by the senior author and led him to abandon it as generally suitable for noncognitive psychological measurement. For a properly constructed set of forced choice scales the sum of a person's scores for all dimensions or traits considered must be the same as the sum of the scores for any other person. This results in what Cattell[2] has called ipsative scores. Scores of this kind can be used to compare a person's standing on a given dimension with his standing on another dimension. They cannot be used to compare his standing on a given dimension with that of another person on the same dimension. Scores suitable for this latter purpose are known as normative scores. The characteristics and limitations of ipsative type scores have been developed in great detail by Clemans[3]. It is true, however, that such scores, even though appropriate only for intra-individual comparisons, can be of great practical value. If they can be freed from the factor of social desirability the forced choice techniques in noncognitive psychological measurement might be justified on this basis alone.

The question still remains whether methods equally economical of construction time and more economical of administration time could be developed which would still tend to suppress the social desirability variable. Now essentially a forced choice set of scores for an individual may be regarded as a set of normative scores expressed as deviations from the person's own mean. The mathematical basis for this method of regarding ipsative scores can be readily demonstrated[3]. One might, therefore, construct sets of items for each postulated dimension and present them in conventional format, scoring them so that a person's score on a dimension is derived from his responses to the individual items for that dimension. The scores for each dimension for a group of subjects could then be appropriately scaled such as by conversion to standard scores. For each person, then, the standardized measures could be expressed as deviations from his own mean of standard scores. Such scores might then again be standardized over persons. This process we will call arithmetical ipsatization.

Suppose now a set of scores on the same dimensions as those just considered had been obtained experimentally by means of a forced choice format and expressed in standard score form. We might regard these scores as comparable to those obtained as indicated in the preceding paragraph. In both cases we would assume that the social desirability factor had been largely suppressed. Although we assume the same items were used in both cases, the administration time would be much less in the former procedure than the latter, since each item was not repeated a number of times with another item.

One objection to the conventional procedure as distinguished from the forced choice procedure is the possibility that most of the variance in the raw scores obtained from the non-forced choice format is due to the social desirability variable. If this is the case the correlation among the dimensions, as measured by the raw scores, will be high, and the subtraction of each person's mean from his own scores

may tend to leave a high proportion of error variance in the residual scores. This in turn would result in low retest reliability of the arithmetically ipsatized scores.

It is therefore reasonable to hypothesize that even though one may suppress the social desirability factor by the arithmetical ipsatization of scores obtained from a format administratively much more economical than the forced choice, nevertheless the former scores might be much less reliable than the latter. This hypothesis can be readily tested by repeating two administrations of both types of formats to the same group of subjects. In the conventional format one must specify the possible responses to each item. These may be simply "yes" or "no" or they may include a series of possible responses varying in degree of agreement, applicability, or what not. When a range of possible responses is provided for each item we may call the format a rating scale.

The Experiment

The forced choice type of noncognitive self-appraisal instrument chosen for this experiment was the Edwards Personal Preference Schedule[4]. It appears to have been carefully constructed and has been the subject of unusually competent research. The EPPS yields scores on 15 different postulated variables. It consists of 225 pairs of items or a total of 450 individual items. Each variable is represented by nine different items. Eight of the nine are used three times each and one of the nine is used four times. The sum of a person's scores for the 15 variables is 210 while the maximum score for any one variable is 28.

A rating scale format was made up by presenting in random order the 135 distinct items in the EPPS. The subject was required to indicate to what extent each statement was a good or poor description of himself by marking one of the following categories on an answer sheet: As good as possible, very good, quite good, rather good, only fair, rather poor, quite poor, very poor, as poor as possible. The possible range of scores for each item was from 1 for "as poor as possible" to 9 for "as good as possible". The raw score for each variable was the sum of the scores for the items assigned to that variable. Thus the highest possible score for a variable was 81.

It should be noted that increasing the range of possible responses to an item tends to increase the reliability of the item scores and hence the score for a variable without increasing the administration time appreciably, if any. As a matter of fact this latitude of response is doubtless less frustrating to the subject than to be forced to choose categorically between "yes" and "no" in a conventional format or between two items in a forced choice format.

The EPPS was administered twice to about 150 undergraduate students with approximately one week between administrations. The conventional format was administered the day after each EPPS administration. Of all subjects tested, 92 had complete data for all four sessions. There were 44 males and 48 females. The average age was 21.9 years. The administration time for the EPPS was approximately one hour whereas for the conventional format it was only about 20 minutes.

For the conventional format the item scores were punched from the answer sheets into IBM cards. Scores for the 15 variables were computed for both administrations on the IBM Type 650 by means of a program developed by Wright[6]. Both sets of rating scale scores were then arithmetically ipsatized to suppress as much as possible of the social desirability factor.

Results

The retest reliabilities of the mathematically ipsatized rating scale scores and those of the EPPS forced choice scores are given in Table 1, together with the differences between them. It will be noted that, contrary to the hypothesis, for only one variable, Abasement, was the reliability higher for the EPPS than for the ipsatized rating scale format and this was only a difference of .01. The greatest difference was for Exhibition. Here the ipsatized rating scale reliability exceeded the EPPS by .24. The average reliability for the ipsatized rating scale form was .87 and

TABLE 1. TEST-RETEST RELIABILITY COEFFICIENTS (N = 92)

Variable	Ipsatized Rating Scale	Edwards PPS	Difference
Achievement	.88	.83	.05
Deference	.78	.67	.11
Order	.90	.81	.09
Exhibition	.88	.64	.24
Autonomy	.89	.76	.13
Affiliation	.91	.80	.11
Intraception	.88	.84	.04
Succorance	.88	.78	.10
Dominance	.91	.82	.09
Abasement	.83	.84	−.01
Nurturance	.84	.81	.03
Change	.85	.80	.05
Endurance	.87	.76	.11
Heterosexuality	.90	.76	.14
Aggression	.87	.72	.15
Mean	.87	.78	.09

that for the EPPS .78, or an average excess of .09 in favor of the rating scale form. This excess in reliability was achieved even though the administration time for the rating scale form was only about one-third that of the EPPS.

SUMMARY

If one may generalize from these data it appears that any advantage which the forced choice type of self-appraisal instrument may have over the arithmetically ipsatized rating scale form must be other than that of greater reliability. Other possible advantages should be investigated in further research.

REFERENCES

1. BAIER, D. E. Reply to Travers' "A critical review of the validity and rationale of the forced-choice technique." *Psychol. Bull.*, 1951, *48*, 421-434.
2. CATTELL, R. B. Psychological measurement: ipsative, normative, and interactive. *Psychol. Rev.*, 1944, 51, 292-303.
3. CLEMANS, W. V. An analytical and empirical examination of some properties of ipsative measures Unpublished doctoral dissertation sponsored in part by the Office of Naval Research, Contract Nonr-477(08), Univer. of Washington Division of Counseling and Testing Services, October, 1956.
4. EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.
5. WRIGHT, C. E. Relations between normative and ipsative measures of personality. Unpublished doctoral dissertation, Univer. of Washington, 1957.
6. WRIGHT, C. E. A program to score rating scales on the IBM Type 650. Technical report, Office of Naval Research Contract Nonr-477(08), Univer. of Washington Division of Counseling and Testing Services, June, 1958.

RELIABILITY OF DIAGNOSTIC JUDGMENTS MADE BY PSYCHOLOGISTS[1]

ALLAN GOLDFARB

City Health Department, Baltimore, Maryland

PROBLEM

The diagnosis is a crucial consideration in deciding upon alternative recommendations for a psychiatric patient. These decisions may involve important consequences for a patient, such as the choice of treatment, commitment to an institution, or separation from military service. One could be more confident of the appropriateness of his course of action if the diagnosis given a patient were the same regardless of which clinician made the evaluation. The purpose of the present study is to investigate both inter- and intrajudge agreement in making diagnostic judgments.

METHOD

Judges. The four psychologists who served as judges were on the staff of the VA Mental Hygiene Clinic in the Baltimore Regional Office. The clinicians were a homogeneous group. They had Ph.D. degrees in psychology and all had completed the VA Clinical Psychology Training Program. Diagnostic evaluations were a routine part of their clinical activities. The Diagnostic and Statistical Manual for Mental Disorders[1] was used by the group as the criterion for classification of the patients. For any particular diagnostic problem the clinicians used fairly similar test batteries, which almost invariably included the Rorschach test. They had in common similar professional experiences, having worked in this clinic from two to five years at the time of this study. This included joint participation in staff meetings and in-service training programs relating to their diagnostic function.

Diagnostic Data. The clinical records were those of veterans who had come to the VA for assessment of service-connected psychiatric disabilities in relation to pension or compensation claims. All had been evaluated by the same psychiatrist, who referred them to the psychology section for diagnostic appraisal. These referrals typically included the more complex and difficult diagnostic cases. The referrals were assigned to the psychologists by the secretary from a roster. These assignments were based on the availability of the clinicians and there was no selectivity by kind of diagnostic problem to be evaluated.

Before seeing a patient the clinician generally had available the referral sheet, which included the psychiatrist's provisional diagnosis, and the patient's military record, which described his psychiatric difficulties in the service. The psychologist's report was focused on the patient's diagnostic evaluation and was rarely more than a page long. The report and the test material comprised the basic data for this study.

Throughout the study the diagnoses given by the psychologists corresponded to the specific psychiatric conditions termed "reactions", as defined in the manual[1]. For every hypothesis tested these specific diagnoses were grouped by the *E* into their relevant "disorder" categories, which consisted of groups of related psychiatric syndromes. Consequently all statistical analyses reported in this study were based on the specific diagnoses grouped in the five categories of the psychophysiologic, psychoneurotic, psychotic, personality, and brain damage disorders. The manner of recording diagnoses[1, p. 50] was reviewed at the start of the study. This was done to insure that the diagnoses given by the judges were consistent in describing the patient's psychiatric conditions in terms of reactions, in contrast to formulating complementary diagnostic data. None of the data were categorized or statistically

[1]The author was a staff member in the VA Mental Hygiene Clinic, Baltimore, Maryland, at the time this study was conducted. Credit is due Dr. Morris Roseman, Chief, Psychological Services, and Dr. James P. O'Connor, VA Consultant, for their encouragement and aid with the design and completion of this study. The statements and conclusions of the author do not necessarily represent those of the VA.

analyzed until the entire study was completed, so as to minimize any bias in the evaluations of the *E*, who served as Judge A. In addition, Judge A completed his evaluations prior to those made by the other clinicians.

Interjudge Agreement

The data consist of diagnostic judgments made by the clinicians on (a) groups of randomly assigned patients; (b) the same group of modified clinical reports; and (c) the individual modified reports.

Procedure and Results. (a) *Groups of Randomly Assigned Patients.* To obtain a sufficient number of cases all evaluations done by the four psychologists between July 1955 and November 1956 were collected from the files. For the group of judges there were from 51 to 71 cases. There were no statistically significant differences in age or education for the four patient subsamples. For each judge the specific diagnoses listed in all his clinical reports were grouped by the *E* under the appropriate diagnostic category. As indicated in Table 1, Chi Square analysis reveals a

Table 1. Classification of Diagnoses Originally Made by Each Judge on His Own Cases*

Diagnostic Category	Judge A N	A %	B N	B %	C N	C %	D N	D %	Total N
Psychophysiologic	8	11	4	6	13	23	2	4	27
Psychoneurotic	34	48	14	21	2	4	18	35	68
Psychotic	9	13	11	16	7	12	11	22	38
Personality	16	22	34	50	35	61	13	25	98
Brain damage	4	6	5	7	0	0	7	14	16
Total	71	100	68	100	57	100	51	100	247

*Chi Square = 62.61; $df = 12$; $p < .01$.

very marked difference in the diagnostic percentage distributions for the judges. These findings were suggestive of considerable interjudge variation beyond chance expectations.

(b) *The Same Group of Modified Clinical Reports.* The diagnostic data consisted of 100 clinical reports. Included were 25 reports from each of the judges' original cases selected by using a table of random numbers. Before the reports were judged they were modified by removing identifying data, which consisted of the patient's name, referral question, diagnosis, and the clinician's name. A member of the staff who did not participate in making the judgments selected the 100 reports and also deleted the specified data from them.

The psychologists were asked to make a diagnostic judgment for each of the 100 reports. This was accomplished in three sessions approximately a week apart. Each clinician made his evaluations independently of the others. All the diagnoses given by each judge were classified by the *E* in the relevant diagnostic category. The results of this procedure are presented in Table 2. Analysis of the data indicated marked interjudge variability beyond chance expectations.

Table 2. Classification and Number of Diagnoses Made by Each Judge on the Same 100 Modified Clinical Reports*

Diagnostic Category	Judge A	B	C	D	Total
Psychophysiologic	13	2	16	3	34
Psychoneurotic	26	30	6	31	93
Psychotic	16	15	14	19	64
Personality	40	47	59	38	184
Brain damage	5	6	5	9	25
Total	100	100	100	100	400

*Chi Square = 43.65; $df = 12$; $p < .01$.

For each judge a comparison was then made of the diagnostic percentage distributions obtained from personal contact with patients (Table 1) and from judging the modified clinical reports with his own original 25 cases excluded. The question raised here is whether or not there is the same kind of significant difference in the diagnostic percentage distributions obtained from the judges, when the evaluations are made from two different types of data. Chi Square analysis indicated a statistically significant change for Judge A only (Chi Square = 10; $p < .05$). Judges B, C, and D consistently assigned approximately the same proportion of cases to each of the diagnostic categories in both situations. Judge A transferred a large group of his cases from the neurotic to the personality category in the second judging situation. The findings indicated that the clinicians assigned significantly different proportions of patients to the various diagnostic categories and, in addition, three of the four judges consistently applied these percentages when dealing with different diagnostic materials.

(c) *The Individual Modified Reports.* Since all four judges had made a diagnostic judgment on the identical 100 modified reports, it was possible to get a measure of interjudge agreement by first comparing all possible pairs of clinicians on the same cases. This was done by counting the number of times any two specified judges agreed by diagnostic category for each of the 100 cases. Agreement between all possible pairs of judges expressed as a percentage was: 63% for A with C; 62% for A with D; 60% for A with B; 58% for C with B; 58% for D with B; and 53% for C with D. The very narrow range of these percentages of agreement suggested that no one judge differed markedly from the other. This indicated that the judges were relatively homogeneous in making these evaluations.

The findings indicated that any two of the clinicians agreed on a diagnosis approximately 60% of the time when evaluating the modified reports. The question is raised as to whether or not this 60% represents a core of cases which would reflect simultaneous agreement by all four judges. In an effort to examine this further, agreement of all four judges on the same 100 cases was studied. The results in terms of the interjudge agreement by diagnostic category of the evaluations made on the same 100 reports were: 32% of the cases reflected unanimous agreement by all four judges; in 40% of the cases there was agreement by three of the four judges, in 15% of the cases two judges agreed on one diagnostic category, while the other two judges agreed on a different diagnostic category; in 12% of the cases two of the judges agreed on a diagnostic category and the other two judges disagreed; and with one case there was complete disagreement. These findings suggest that 60% of the cases does not represent the minimum number with which there would be agreement between all four judges.

Intrajudge Agreement

The data include repeated diagnostic judgments made by each psychologist on (a) clinical reports; and (b) test data which he had initially completed.

Procedure and Results. (a) *Clinical Reports.* Intrajudge repeatability was determined by counting the number of perfect agreements between the diagnoses grouped by categories made by each of the judges on his own original 25 reports and the same reports with identifying data deleted, and dividing by 25. The percentage of agreement between the repeated evaluations for each judge was: 68% for Judge D; 64% for Judge A; 64% for Judge B; and 52% for Judge C. There was no statistically significant difference in the number of perfect agreements for the judges as measured by the Chi Square test. The average agreement for all the judges was approximately 60%. These findings suggested that each judge was equally consistent in repeating his own evaluations.

(b) *Test Data.* Each clinician was presented with the test data of 25 cases he had worked-up originally and was asked to make a diagnosis. Approximately two weeks later each judge was asked to repeat his evaluation of the same test data. The

judges made their evaluations independently of the others. The percentage of agreement between the repeated judgments grouped by categories for each clinician on the same test data was: 64% for Judge A; 60% for Judge B; 60% for Judge C; and 56% for Judge D. Inspection of these percentages suggests that there are no striking differences among the judges. The average agreement was 60%. One concludes that the judges were equally consistent in repeating their evaluations based on test data. It is to be noted that these findings are about the same as those determined in the situation where the judgments were made solely on the clinical reports.

Discussion

The judges differed significantly in the proportions of their diagnostic evaluations grouped by categories, and in addition, tended to consistently apply these proportions. This finding may be explained in several ways: (a) the clinicians may have consistently focused on different facets, symptoms, or levels of patient behavior in deriving a diagnosis according to their theoretical persuasion and experience, *i.e.*, there is no one 'right' diagnosis; (b) they may have evaluated the same diagnostic features, but repeatedly used the diagnostic nomenclature differently; and (c) the judges may have had preconceived proportions by which they typically grouped the patient population.

However, when one considers that inter- and intrajudge agreement was on the order of 60%, one may question the stability of the diagnostic criteria used by the clinicians in making their evaluations.

In assessing the reliability findings derived from the psychologists' evaluations of the clinical reports which had identifying information deleted, and the test data, a crucial consideration is that the judges did not have face-to-face contact with patients. It should be noted that both these research situations are essentially artificial in that clinicians do not make a diagnosis in this manner in their clinical practice. This leaves unanswered the question of the possible influence that direct contact with patients might have had on the reliability of these judgments made by the clinicians. However, the finding of a significant difference among the clinicians in the proportions of their diagnostic evaluations grouped by categories, the clinical data having been obtained by direct contact with patients, suggests the possibility that the marked variation in the diagnostic judgments derived from both the modified clinical reports and the test data might not have been materially lessened if the judges had also interviewed the patients. The results of this study indicate that a satisfactory degree of reliability of diagnostic judgments was not consistently obtained in either the realistic clinical situation or in the basically artificial ones.

The seeming discrepancy between the significant differences in the proportions of cases classified by diagnostic category and judge repeatability of the magnitude of 60% is very likely a reflection of what might be termed the classification of core versus non-core cases. A core case can be considered one in which the judges agree on its diagnosis; the opposite situation would prevail in varying degrees with the non-core cases. It is very probable that with certain clinical manifestations there will be complete judge agreement on the diagnosis, while with other kinds of cases the judges would be more prone to disagree on the diagnosis. An example of the former situation might involve classifying the patient who is delusional and hallucinating, while in the latter situation a patient who presents a less striking symptomatic picture would more probably be diagnosed differently by the clinicians. The data in Table 1 support this hypothesis, since marked diagnostic differences among the judges occur in the personality and the psychoneurotic classifications. In terms of this possible explanation the apparent discrepancy is likely due to the judges' having agreed on the core cases, but the non-core cases were proportionately large enough in number and diagnosed sufficiently different as to significantly influence the statistical distribution of all the cases evaluated. As suggested by the data derived from the judges' evaluations of the same 100 modified clinical reports, the core of cases agreed upon by the four judges is considerably less than 60% of all the cases.

This study attempts to assess the reliability of the final diagnostic judgments of clinicians. It would seem appropriate in terms of future research to study those factors which initially influence diagnostic decision-making by psychologists. The factor of direct patient contact would seem particularly crucial to evaluate further. It would also be of value to clarify the characteristics of core and non-core cases.

The fundamental question raised in this study was whether or not the diagnoses given to groups of patients depended to a significant degree upon which clinician made the evaluation. The results suggest that the answer to this question is substantially "yes."

Summary

This study attempted to evaluate inter- and intrajudge agreement in making diagnostic judgments. The two general hypotheses investigated were: (a) the judges agree in their diagnostic evaluations; and (b) each judge is consistent in making repeated diagnostic evaluations. Experimental procedures to test these hypotheses were described. The results indicated that there was significant variability among the judges in the proportions of their diagnostic evaluations grouped by categories. The clinicians also tended to apply these proportions in a consistent manner. Interjudge agreement based on evaluating modified clinical reports was on the order of 60%. Intrajudge agreement based on making repeated evaluations of the same test data was also approximately 60%. The core of cases with which there is unanimous agreement by the judges is apparently considerably less than 60% of all the cases of the type evaluated in this study. The overall finding of this study suggests that the diagnoses given to groups of patients are in large measure dependent on which clinician makes the evaluations. Suggestions for future research are indicated.

References

1. American Psychiatric Association, Committee on Nomenclature and Statistics. *Diagnostic and statistical manual for mental disorders.* Washington: American Psychiatric Association, Mental Hospital Service, 1952.

THE VALIDITY OF CLINICAL JUDGMENTS OF SCHIZOPHRENIC PATHOLOGY BASED ON VERBAL RESPONSES TO INTELLIGENCE TEST ITEMS[1]

NELSON F. JONES, JR.

Northwestern University

The reliability of clinical judgment has been the subject of a series of studies using single responses to intelligence test items obtained from psychiatric patients, and judgments by trained clinicians based on a seven-point rating scale. It has been possible to demonstrate reliability in judgments of disorganization in schizophrenic thinking [2, 3, 4], as well as more specific aspects of schizophrenic thinking [5, 6].

The question now arises whether clinicians can make global appraisals of complete test protocols with as great reliability as they can judge single test items. This paper extends the study of reliability, defined as interjudge agreement, to global

[1]This study is part of a larger project being conducted at Northwestern University under the direction of Professor William A. Hunt through contract 7onr-450(11) with the Office of Naval Research. Facilities for collecting the experimental materials and establishing the criterion were provided by the VA Hospital, Hines, Illinois. The opinions expressed, however, are those of the author and do not represent opinions or policy of the Naval service nor of the Veterans Administration. The study was submitted to the Graduate School of Northwestern University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to thank Dr. W. A. Hunt, Dr. R. I. Watson, Dr. E. L. Clark, and Dr. Janet A. Taylor for their guidance and assistance.

judgments in which whole protocols are ranked according to severity of schizophrenic pathology. It also introduces a validity measure. The validity of the judgments made here is tested by measuring the extent to which they agree with an external criterion. In addition, the effect upon judgment of varying the amount of test material presented to the judges was studied, as was the effect of varying amounts of training and experience. The specific hypotheses investigated were:

1. The judges, both experienced clinicians and naive undergraduates will demonstrate a high degree of inter-judge agreement in ordering the protocols, but that of the trained clinicians will be greater under all conditions.

2. Both sets of judges will demonstrate a better than chance ability to order the protocols according to the external criterion of severity of pathology. Further, the agreement of the trained clinicians with the external criterion will be closer than that of the naive undergraduates.

3. Increasing the amount of test material to be judged within the limits of the two sub-tests under consideration will result in decreased reliability and validity for both groups.

Method

Establishment of the Criterion. A group of twelve neuropsychiatric patients was selected by the author in consultation with a member of the Clinical Psychology Service of the hospital from which the patients came. The group was chosen from male patients whose records indicated no other psychiatric diagnosis than schizophrenia and was selected to represent a wide range of pathology.

A criterion group of three clinical psychologists, whose selection will be discussed later, observed through a one-way viewing screen an interview with each of the twelve patients conducted by a fourth staff psychologist. The interviews were heard by the observers through the monitoring system of a tape recorder and were simultaneously recorded. The interviews were semistructured and covered at some time each of four general points decided upon previously by the three observers as relevant to the decisions which they were to make.

Each of the criterion judges then reviewed all available case history material and projective tests results, including a Rorschach and Sentence-Completion or TAT. No attempt was made to hide the identity of the patients at any time. After both the interview and the review of written material, each judge was asked if he felt the patient was other than schizophrenic, and then, to make a rating of how severe he judged the patient's schizophrenic pathology to be. Ratings were made on a seven-point scale of severity of schizophrenic pathology on which point one represented minimal pathology and point seven represented maximal pathology. Each judge was then asked to make an independent ranking of all twelve patients according to severity of pathology.

After these independent judgments had been recorded, the three judges met in conference and, as a group, ranked all twelve patients. They then selected from the twelve cases seven which they felt they could most clearly and unambiguously rank according to severity of schizophrenic pathology exhibited at the time of the interview. The ranking of these seven patients became the experimental criterion.

Immediately after the interview, each patient was given the Vocabulary and Comprehension sub-tests of the Wechsler Adult Intelligence Scale. The verbatim test responses of the seven patients selected became the experimental materials for this study. All responses of each patient to each sub-test were mimeographed on separate sheets, and in addition, for the condition in which only half the Vocabulary responses were used, odd-numbered Vocabulary items were mimeographed on separate sheets, making three sub-test protocols for each patient.

Subjects. Three groups of judges comprised the subjects in the study. The first group, the three criterion judges, each held the Ph.D. degree in Clinical Psychology with clinical experience ranging from nine to fourteen years. Each judge had participated in our previous studies of the reliability of clinical judgments and had demonstrated a high degree of reliability with the other criterion judges. The second group, the experimental judges, comprised forty-eight experienced clinical psychologists used in our previous studies. Each possessed the Ph.D. degree and at least four years of professional experience. The third group, also experimental judges, contained forty-eight undergraduate students from psychology classes at Northwestern University, none of whom had participated in the previous studies or had experience with psychiatric patients.

Procedure. The experimental materials were distributed by mail to the clinical psychologists acting as experimental judges and were presented to the student judges in a classroom situation. Each of these two groups was subdivided into three groups of sixteen persons each, and each of the sub-groups was presented with one of the following combinations of material: (a) seven protocols consisting of only one-half the vocabulary responses of each patient, (b) a complete vocabulary protocol of each patient, or (c) a complete vocabulary protocol and a complete comprehension protocol for each patient.

The following instructions were given to all experimental judges:

> We are going to present you with a number of sub-test protocols taken from intelligence scales. One of the ways in which the pathology of schizophrenia may express itself is through disordered thinking which results in atypical, deviant or "abnormal" responses to the items on such a test. The qualitative interpretation by the clinician of such test responses is one of the bases upon which he makes a clinical or diagnostic interpretation. All the individuals who contributed these protocols were diagnosed as schizophrenic. None of the cases was complicated by mental deficiency or known organic conditions. The extent of the pathology exhibited, however, is not uniform. In some cases it is less extreme than in others.
>
> We are going to ask you to rank these protocols according to the severity of the pathology exhibited in the responses and to indicate on the attached form the order in which you rank them. In making these judgments, we are asking you to concentrate upon the severity of the pathology exhibited in the responses. We are not interested in such things as prognosis, chronicity, therapeutic indications, or sub-type designation. In essence, what we are asking you to do is judge how "schizophrenic" each of these records is in relation to the others presented with it.

Inter-judge reliability within each condition was determined by correlating each judge's ranking with the mean ranking of the other fifteen members of his group, using Spearman's *Rho*. Validity coefficients for each judge were determined by comparing his ranking to the criterion ranking, again using the Spearman *Rho*.

Results

Criterion selection. Although each criterion judge was acquainted to a greater extent with some patients than with others prior to the time at which the criterion was established, a fact which would be expected to lower agreement on the ratings, inter-judge reliability was quite high. Table 1 summarizes the intercorrelations between judges in establishing the criterion. Ratings based on the interviews and on the test and history materials were absolute judgments based on a seven-point scale and were compared by the Pearson's r correlation. Independent rankings of the twelve patients were compared using Spearman's *Rho*, as were each judge's rankings and the conference results. From the ranking of the twelve patients established in the conference, the judges chose the patients ranked "1", "3", "5", "7", "8", "10", and "12" as most clearly discriminable.

Table 1. Summary of Criterion Correlations

Material Correlated and Statistic Used	Judges 1 and 2	1 and 3	2 and 3
Interview Ratings: r on 12 patients	.775**	.553*	.841**
Test-History Ratings: r on 12 patients	.827**	.721**	.906**
Individual Rankings: *Rho* on 12 patients	.853**	.587*	.776**
	1	2	3
Agreement of Each Judge with Conference Ranking: *Rho* on 12 patients	.916**	.948**	.832**

*Significant at the .05 level of confidence
**Significant at the .01 level of confidence

Six months later a follow-up was made of the hospital status of these patients. Their status then is listed below in order of their ranking. "1" Discharged, maximum hospital benefits; "3" Discharged on trial visit; "5" Discharged on trial visit, in out-patient treatment; "7" Transferred for long-term hospitalization elsewhere; "8" Transferred for long-term hospitalization elsewhere; "10" Discharged on trial visit; "12" Transferred for long-term hospitalization elsewhere.

Experimental results. Using a z-score transformation of the individual reliability and validity *Rhos*, a mean correlation was determined for each of the six experimental groups for each measure. This information appears in Table 2. It is clear that the mean reliability of even the poorest group is quite high, indicating that the task can

TABLE 2. MEAN RELIABILITY AND VALIDITY COEFFICIENTS FOR EACH EXPERIMENTAL GROUP OF 16 JUDGES

Condition	Partial Vocabulary	Whole Vocabulary	Whole Vocabulary and Comprehension
Clinicians			
Reliability	.955	.926	.862
Validity	.686	.611	.711
Undergraduates			
Reliability	.834	.867	.762
Validity	.446	.430	.485

be accomplished reliably even by the group of inexperienced judges who received the largest amount of material, while the group of experienced clinicians using the smallest amount of material agree almost entirely in their judgments. These results agree with those discussed by Hamlin [1].

For validity, we find evidence almost as clear as that for reliability. The significance of the validity measures is indicated further by the fact that all forty-eight of the forty-eight clinicians' individual validity coefficients fell in the predicted direction of a positive correlation with the criterion, and forty-two of the forty-eight undergraduates' correlations were also positive. Twenty-one clinicians exceeded the .05 level of confidence in their ratings, and nine undergraduates did so. While experience and training make the clinician a better judge, it is obvious that this task is being performed at a level more accurate than would be expected by chance.

Differences in reliability and validity attributable to experience and training or to the amount of material used in making these judgments were tested for significance by using the individual reliability and validity coefficients, as converted to z-scores, in analyses of variance. The six combinations of experience and amount of material presented were the conditions compared. Hartley's test for homogeniety of variance gave no indication that analysis of variance was inappropriate.

TABLE 3. I. ANALYSIS OF VARIANCE OF RELIABILITY COEFFICIENTS

Source of Variation	Sum of Squares	df	Mean Square	F
Experience	50,738	1	50,738	21.425**
Amount of Material	34,642	2	17,321	7.313**
Interaction	6,445	2	3,222	1.361
Between Groups	91,825	5	18,365	7.755**
Within Groups	213,139	90	2,368	
Total	304,964	95		

II. ANALYSIS OF VARIANCE OF VALIDITY COEFFICIENTS

Source of Variation	Sum of Squares	df	Mean Square	F
Experience	25,058	1	25,058	20.547**
Amount of Material	2,409	2	1,204	.988
Interaction	568	2	284	.233
Between Groups	28,035	5	5,607	4.598**
Within Groups	109,761	90	1,219	
Total	137,796	95		

**Significant at the .01 level of confidence

The difference in performance between clinicians and undergraduates indicates a clearly significant increase in reliability as a result of increased training and experience. The increase in amount of material, however, results in a significant decrease in reliability. The non-significant interaction effect between experience and amount of material indicates that, contrary to expectation, experienced clinicians are not relatively better than undergraduates at agreeing with each other when increased amounts of material are presented to them.

In the case of validity, the clinicians again excel over the undergraduates. However, increasing the amount of material to be judged did not produce a significant change, the interaction effect is again non-significant.

It was possible to separate the clinical judges by level of previous experience on our prior rating studies and to test by analysis of variance, the differences between the three levels of prior experience. No significant differences in either validity or reliability were found which could be attributable to differing amounts of experience with the previous tasks, showing such practice effects to be less important here than in a previous study of context effects [8].

While both clinicians and undergraduates agreed with the criterion at a better than chance level, none of the groups' mean judgments agreed exactly with the criterion ranking of the patients. Distribution of the percentages of agreement between the 112 individual judgments made by each group and the criterion ranking, while another way of presenting the same trend shown in Table 2, indicated that as the magnitude of error in a single case increases, the likelihood of that error being made rapidly decreases in this sample of judgments. This, of course, indicates a high probability of accuracy in the clinical situation where single judgments are required.

Summary

Forty-eight trained clinicians and forty-eight undergraduates ranked three amounts of Vocabulary and Comprehension sub-test responses according to how schizophrenic they felt the person contributing the protocol to be. Interjudge reliability and correspondence to an external validity measure were determined. Both clinicians and undergraduates were more reliable and more valid than chance expectations. The clinicians were significantly more reliable and closer to the validating criterion than were the undergraduates. As the amount of material presented to be judged increased, reliability decreased for both clinicians and undergraduates, but validity was not affected. Clinicians were relatively no better at handling increased amounts of material than were the undergraduates.

References

1. Hamlin, R. M. The clinician as judge: implications of a series of studies. *J. consult. Psychol.*, 1954, *18*, 233-238.
2. Hunt, W. A. An actuarial approach to clinical judgment. In Bass, B. (Ed). *Objective approaches to personality assessment.* New York: Van Nostrand. In press.
3. Hunt, W. A., and Arnhoff, F. N. Some standardized scales for disorganization in schizophrenic thinking. *J. consult. Psychol.*, 1955, *19*, 171-174.
4. Hunt, W. A., and Arnhoff, F. N. The repeat reliability of clinical judgments of test responses. *J. clin. Psychol.*, 1956, *12*, 289-290.
5. Hunt, W. A., and Jones, N. F. The reliability of clinical judgments of asocial tendency. *J. clin. Psychol.*, 1958, *14*, 233-235.
6. Hunt, W. A., and Jones, N. F. Clinical judgment of some aspects of schizophrenic thinking. *J. clin. Psychol.*, 1958, *14*, 235-239.
7. Hunt, W. A., Jones, N. F., and Hunt, Edna B. Reliability of clinical judgment as a function of clinical experience. *J. clin. Psychol.*, 1957, *13*, 377-378.
8. Jones, N. F. Context effects in judgment as a function of experience. *J. clin. Psychol.*, 1957, *13*, 379-382.

A COMPARISON OF RANDOM AND JUDGMENTAL METHODS OF DETERMINING MODE OF OUTPATIENT MENTAL HYGIENE TREATMENT

FLOYD K. GARETZ[1], RICHARD C. KOGL[2], AND DANIEL N. WIENER[3]

Veterans Administration, St. Paul, and University of Minnesota

Problem

Much has been written about ways of deciding how to treat patients with emotional problems. The best way of arriving at a clinical judgment for treatment is presumably provided by the "team", consisting of psychiatrist, psychologist, and social worker, each contributing his specialized knowledge to the decision[1, 2, 3, 4]. In a recent study, Wiener and Raths[5] questioned the efficacy of clinical judgment in deciding upon the best form of psychotherapy. They found that each of the three disciplines on the team—psychiatry, psychology, and social work—was about equally accurate in predicting what the therapist would, at the conclusion of treatment, describe as the preferred method of therapy. No group was very good, however, nor did meeting together as a team improve the predictions significantly.

To us, the usefulness of clinical judgment as a way of deciding on the form of psychotherapy was strongly suspect. We decided to try to compare its relationship to success of treatment, with the success following assignment of cases to a particular therapy randomly.

Method

Twenty-seven male white veterans applying for treatment at a mental hygiene clinic were assigned either to a "clinical judgment" group (Group 1), or a "random number" group (Group II). In Group I, 13 patients were prescribed and received drug therapy or psychotherapy as determined by the best judgment of the clinic team and the psychotherapist. The 14 patients in Group II were assigned to either drug therapy, or psychotherapy, from a random number table, and the therapist was bound to this "artificial" decision.

Two psychiatrists treated all the patients. They chose drugs at their own discretion, and saw "drug therapy" patients for no more than 15 minutes each, twice a month. Psychotherapy patients were seen weekly for 30 to 50 minutes. After three months of treatment, the therapists rated all patients on a five-point scale of treatment success ranging from "worse" to markedly improved."

Results

The data were analyzed in a fourfold table: 14 randomly-assigned cases, and 13 cases assigned through clinical judgment, each subdivided into "improved," and "unimproved" groups. "Unimproved" included the "unchanged" and "worse", and "improved" included those showing any degree of improvement. Three cases of each group were unimproved, 11 and 10 of the "random" and "clinical judgment" groups respectively were improved. The difference was not significant. Because of the small number of patients, we were not able to subdivide our results (as we had planned) according to form of treatment. Nor could we readily generalize from such sparse data.

Summary and Conclusions

Our initial hypothesis was that current methods attempting to match patient and mode of outpatient therapy by clinical judgment are of questionable validity. We studied the results of assigning patients to therapy randomly, as compared with clinical judgment, and found the improved rates to be not significantly different. We

[1]Psychiatric resident, VA Hospital, Minneapolis, Minn.
[2]Psychiatrist, private practice, Minneapolis and VA Mental Hygiene Clinic, St. Paul, Minn.
[3]Chief Clinical Psychologist, VA Mental Hygiene Clinic, St. Paul and University of Minnesota.

concluded from our data, considered with the other results cited, that more caution should be exercised in current methods of deciding clinically which patient should have any specified form of outpatient therapy.

References

1. Ackerley, S. The clinic team. *Amer. J. orthopsychiat.*, 1947, *17*, 191-195.
2. Drew, A., Jr. Teamwork and total patient care. *J. Psychiat. Soc. Work.*, 1953, *23*, 25-31.
3. Healy, W. and Bronner, A. F. Orthopsychiatry: an overview. *Amer. J. orthopsychiat.*, 1955, *25*, 472-474.
4. Lemkan, P. V. Mental hygiene in public health. *Public Health Reports*, Wash., 1947, 62, 1151-1162.
5. Wiener, D. N. and Raths, O. N., Jr. Contributions of the mental hygiene clinic team to clinic decisions. *Amer. J. orthopsychiat.*, 1959 (April).

A COMPARISON OF PARENTAL AND SELF-EVALUATIONS OF PSYCHOPATHOLOGY IN CHILDREN[1]

EUGENE E. LEVITT

Indiana University Medical Center

The mother of a child guidance clinic patient is usually considered to be a primary source of information about the child, even though it is thought that they frequently have a distorted perception of their child. Nevertheless, the mother's statement must often be accepted as valid in the absence of contradictory information. It is a sort of "best estimate" of the child when no other estimate is available. Despite the prevalence of the belief in the fallibility of the mother's perception, there have been few experimental efforts to check it. The job of rounding up estimates of the child other than that of the mother is a tedious and time consuming one. Even if it could be accomplished reliably, there is always the possibility that differences in the perception of the child by the mother or teacher are a function of actual differences in the child's behavior in two different settings. By the time the clinician obtains sufficient data through observation of the child to check the mother's report, the contacts with the clinic may have changed the child appreciably.

Helper (2), in one of the few studies bearing on the question, found low, positive correlations between self-evaluations of a group of 8th and 9th grade children and evaluations of them by their parents.

This experiment attempts to check the mother's perception of the emotionally disturbed child.

Subjects and Procedure

The Children's Manifest Anxiety Scale (CMAS) (1) was administered to 73 consecutive children between the ages of 8 and 13, who were seen diagnostically at a large community child guidance clinic. In view of the high incidence of reading deficiencies among the clinic's patients, each item in the scale was read to the child by the psychologist and his responses recorded.[2] Simultaneously, an identical form of the test, except that the pronouns "I", "my", etc., were replaced by "my child", "his", etc., was administered by the social worker to the mother of each child. The mothers were told that they could omit an item if they lacked sufficient information with which to respond to it. In this way, the child and his or her mother described the child's symptoms by means of an identical questionnaire.

[1]The experimental work summarized in this paper was supported by the Institute for Juvenile Research of Chicago, an agency of the Department of Public Welfare of the State of Illinois. It was read at the 1959 meeting of the Society for Research in Child Development in Bethesda, Md.

[2]Wirt and Broen (3) found that individual and group administrations of the CMAS yield similar results.

Results

The mean score on the 42 anxiety items (the A-Scale) for the children's responses was 20.82 with a *SD* of 8.14, practically identical with the mean of 20.96 and *SD* of 7.90 obtained for an earlier sample of a hundred cases from the same clinic [4], but significantly higher than the means reported for several different samples of normal children of comparable ages [3].

The mean score for the mothers was 18.82, *SD* 7.47. The difference between means, analyzed by an appropriate *t*-test for correlated means, is not significant. Furthermore, if we prorate the mother's scores for omitted items, the mean becomes 20.19 with a *SD* of 7.76, practically identical with that of the children. A similar lack of difference prevails for the 11 Lie Scale items. The mean for the children was 4.07, *SD* 2.27. For the mothers, the mean was 3.30, *SD* 2.44. The *t* of 1.91 is not significant. Prorating for omitted items, the mean for the mothers becomes 3.44 with a *SD* of 2.39. We see, then, that on the average, the children and their mothers perceived the children as disturbed to the same degree. However, the correlation between scores for pairs of mothers and children on the A-Scale unprorated is .14, prorated, .09. The respective figures for the Lie Scale are −.07 and −.02. None of these coefficients is significant.

Considering the degree of agreement item by item for all items between pairs of mothers and children, not counting items omitted by the mother, we find that the agreement is 56.69% which does not differ significantly from the chance expectancy of 50% ($t = 1.15$). The agreement on the A-Scale items alone is 57.53%, on the Lie Scale items, 56.46%, a nonsignificant difference.

If we count one for each item agreement and zero for each non-agreement and then compute an items × subject analysis of variance, we find significant *F*-ratios for items on both the A-Scale and the Lie Scale (the respective *F*s are 4.23 and 2.69, both significant at the .01 level). This suggests the importance of noting those items on which there was a greater than chance agreement. There were 10 such items on the A-Scale and three on the Lie Scale. The agreement on these 13 items ranges from 62 to 73%. The 10 A-Scale items do not seem to cluster in any logical fashion.[3] Four seem to reflect inner states, two are somatic symptoms, three are other observable states, and one concerns school work.

Discussion

Children and their mothers agree on the average degree of psychopathology among the children, and this average agrees with what has been found previously for disturbed children. However, the lack of correlation between the scores of mothers and their children, and the lack of item by item agreement, means that the individual mother on the average perceived her child's symptoms differently than he did. Since the clinic is usually concerned with the nature of symptoms, we should attempt to determine which one is perceiving realistically, mother or child.

The implicit assumption in the self-administered adjustment inventory or similar device is that an individual knows his own state better than any outside

[3]The A-Scale items were:

7. At times (I) (my child) feels like shouting. (65% agreement)
8. At times (I) (my child) wish (I) (he) could be very far from here. (65%)
13. (I) (he) feels alone even when there are people around (me) (him). (73%)
15. (I) (he) get nervous when things do not go the right way for (me) (him). (63%)
19. Often (I) (he) have trouble getting (my) (his) breath. (64%)
23. (I) (he) have to go to the toilet more than most people. (67%)
26. (I) (he) have trouble swallowing. (72%)
37. (I) (he) often get lonesome when (I) (he) am with people. (69%)
39. (I) (he) am afraid of the dark. (63%)
40. It is hard for (me) (him) to keep (my) (his) mind on (my) (his) school work. (62%)

The L-Scale items were:

10. (I) (he) would rather win than lose in a game. (63%)
30. (I) (he) am always good. (70%)
36. (I) (he) tells the truth every single time. (65%)

observer. Looking at it this way, we must give the child the benefit of the doubt and assume that his is the clearer perception. Furthermore, Rynerson[4] found a positive correlation between the child's perception of the severity of his symptoms based on the CMAS, and estimates of severity by the examining psychiatrist in another group of patients from the same clinic.

Examination of the protocols suggested two possible explanations for the lack of correspondence of pairs of scores. It is reasonable to suppose that the symptoms of children who are labelled as acting-out, *i.e.*, those whose pathological behavior is more manifest, will be better known to the parents than those of children who do not act out. In our sample, there were 15 children labelled as acting-out problem cases by the psychiatrist. The mean percentage of agreement between mothers and children in this group is 59.01 which does not differ significantly from the percentage for the remaining 58 cases (a median test chi-square is 0.43.) We find, however, that the parents of five of these 15 children had better than chance agreement with the child. This is 33.3% as opposed to 11.9% for the remaining 58 cases and 16.4% for the total N of 73. The difference in the number of cases in which there is a greater than chance agreement between the acting-out cases and the other cases is significant, yielding a probability of .05 by Fisher's exact method. We must therefore consider that one possible explanation of the lack of agreement on symptoms is simply that the parents are unaware of many of the symptoms because they are not manifest.

A second possibility was that the lack of agreement was a function of excessive guessing concerning the child's symptoms by the parent. This hypothesis was tested in the following manner. There were 14 mothers who omitted seven items or more and 37 mothers who did not omit any items. The average percentage of agreement with the child's responses for the former group was 61.28 and 55.63 for the latter group. The difference is not significant when tested parametrically ($t = 1.68$) but is significant with a nonparametric test (a median test chi-square is 5.87, $p < .05$). Further scrutiny of the protocols indicated that the discrepancy between the parametric and nonparametric tests was a function of a single, unusually low degree of agreement (26.2%) in the group of mothers who had omitted seven or more items. If this score is eliminated, the mean agreement for that group rises to 63.98%, the pooled variance is reduced by 15%, and the t is now 2.63, $p < .05$. We conclude that this is another possible explanation of the lack of agreement between the responses and scores of mothers and children.

Summary

The purpose of this study was to attempt to assess the perception of the symptoms of an emotionally disturbed child by the mother. The CMAS was administered to 73 clinic patients and also filled out for them by their mothers. There was no difference between mean scores for mothers and children, and both were higher than means obtained from normal samples. There was, however, no correlation between pairs of mother-child scores, and item-by-item agreement between pairs was not greater than chance. Some evidence was presented which suggests that mothers may be unaware of their children's symptoms.

References

1. Castaneda, A., McCandless, B. M. and Palermo, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpmt.*, 1956, *27*, 317-337.
2. Helper, M. M. Parental evaluations of children and children's self-evaluations. *J. abnorm. soc. Psychol.*, 1958, *56*, 190-194.
3. Levitt, E. E. Ecological differences in performance on the Children's Manifest Anxiety Scale. *Psychol. Rep.*, 1957, *3*, 281-286.
4. Rynerson, Mary N. The relationship between scores on the Children's Manifest Anxiety Scale and clinical judgments of anxiety. Unpublished master's thesis, Univ. of Chicago, 1957.
5. Wirt, R. D. and Broen, W. E. The relation of the Children's Manifest Anxiety Scale to the concept of anxiety used in the clinic. *J. consult. Psychol.*, 1956, *20*, 482.

STANDARDIZATION OF THE CHILDREN'S PICTURE INFORMATION TEST[1]

KATE L. KOGAN AND RICHARD L. CRAGER

Department of Psychiatry, University of Washington School of Medicine

University of Washington

PROBLEM

The need for a relatively rapid technique of assessing intellectual development in young children with serious motor handicaps had led to the development of the Children's Picture Information Test (CPIT).[2] The test was designed for use with the population of the Spastic Children's Clinic and Preschool in Seattle, Washington. In a preliminary study of 50 young cerebral palsied children the correlation between CPIT and standard psychometric measures was .82 (5). These findings made it apparent that further refinement of the instrument and its standardization on a non-handicapped population was warranted.

METHOD

The Test. Several minor modifications were made in the test material before the standardization procedures were initiated. The original form had consisted of 34 sets of items; six additional sets of items brought the total to 40. Originally the several pictures on a single page were mounted on different colors; in the standardization form all pictures on a single page were mounted on the same color, so that color preference could not possibly influence the child's selection of a response.

The method employs a multiple choice task requiring the child to indicate the one picture (of four possibilities) which belongs with each key picture. An effort was made to restrict the content largely to everyday household situations so that the immobilized child of confined environment would have had a chance for familiarity with the majority of the objects pictured. The sets of pictures are mounted on large pages in book form. Colored pictures were selected from picture dictionaries and other object picture books designed for young children.

SUBJECTS[3]

Standardization. The experimental design called for administering the test to 50 children in each of eight six-month age intervals between 2 years 0 months and 5 years 11 months, making a total of 400 children. It was recognized that there was no easy source of a heterogeneous population sample at this age level, comparable, for example, to using classroom groups of older children. Therefore, as partial insurance against biased selection, the children in each age group were classified according to their parents' occupations so that the distribution would conform to the distribution of occupations according to 1950 US census figures. Attempt was also made to balance the groups as evenly as possible for numbers of boys and girls, and for age in months within the six month range.

Since the majority of subjects were made available through various kinds of nursery school groups, it was necessary to examine all children in any given group who expressed a desire to participate. In this way a reservoir of data on 568 children

[1]This study was initially supported by a grant from the Dean's Cerebral Palsy Fund, University of Washington School of Medicine, and was completed under a research grant #B-1611 from the National Institute of Neurological Diseases and Blindness, Public Health Service.

[2]The CPIT test material is assembled by hand by the Spastic Aid Council, 1850 Boyer Avenue, Seattle, Washington. It will be provided to professional workers at a nominal charge to cover costs of assembling and postage.

[3]The investigators are indebted to the Seattle Day Nursery Association, the Family Life Division of the Seattle Public Schools, and the W. C. Bryant Elementary School for generously making their extensive facilities available; the YWCA Nursery, University Preschool and Kindergarten, Carousel Nursery School, Southtown Preschool, Gingerbread House, and Lockhaven Nursery also gave invaluable cooperation.

was accumulated. Within each age group the children's records were arranged in the chronological order in which they were tested; the first fifty children who best fitted the ideal distributions of sex, age in months, and parents' occupation were selected for the standardization population. The characteristics of the subjects included in the standardization are detailed in Table A of the Appendix.[4] The occupational distributions obtained in the various age subgroups were not significantly different from the model.

Validity Group: Correlation with Stanford Binet. A group of 50 subjects was selected from the total reservoir available so that their distribution of ages from 2-0 to 6-0 was balanced and so that the occupation of the parents approximated the census groupings. Each subject was given the Revised Stanford Binet, Form L, in addition to the CPIT. For practical reasons it was necessary to use only children whose parents could make them available at a time other than a regular nursery school session, for example, at their own home by appointment, or at a neighbor's home. This was, then, more strictly a volunteer population than the larger standardization population, and subject to the distortions inherent in sampling volunteers. The mean IQ for the group was 109.1 with a range from 83 to 142, and a standard deviation of 13.7. The detailed composition of the group is given in Table B of the Appendix. A Chi-square test for both occupational deviation from the US census model, and for equal distribution of ages throughout the range yielded no significant deviations. The sex distribution was somewhat less than adequate, with 35 males and 15 females in the group.

Handicapped Group: Correlation with Stanford Binet. The test material had been supplied on an experimental basis to a number of other professional workers in this country and abroad for experimental use in clinical situations. The senior author had also continued to use the instrument with the clinic population at the Spastic Children's Clinic and Preschool in Seattle, Washington. From these sources the records of 59 children were selected to whom both the Revised Stanford Binet and the CPIT had been successfully administered.

Test-Retest Reliability Group. A group of 50 subjects was selected from the total reservoir available, in the same manner as described above for the validity group. Each subject was given a second CPIT at an interval of 1-3 weeks from the date of his first test. Occupational and age distributions were not significantly deviant from the model. There were 26 males and 24 females. The group statistics are given in Table C of the Appendix.

Results

Scores by Age Intervals. The total raw score (number of correct responses) for each subject was tabulated, and the data analyzed by age-interval groups. The findings are summarized in Table 1. Observation indicated that both mean and median

Table 1. Distributions of Scores by Age Intervals

Age group	Median	Mean	Standard Deviation
2-0 to 2-6	7	7.3	5.2
2-6 to 3-0	10	10.1	5.6
3-0 to 3-6	18	18.3	7.6
3-6 to 4-0	23	22.7	7.9
4-0 to 4-6	30	29.4	6.4
4-6 to 5-0	34	31.3	8.3
5-0 to 5-6	37	36.0	4.4
5-6 to 6-0	38	37.2	2.4

[4]An Appendix giving more detailed data about subjects and results has been deposited with the American Documentation Institute. Order Document No. 5914, remitting $1.25 for 35-mm. microfilm, or $1.25 for 6 x 8 inch photocopies.

scores increased from each age group to the adjacent higher age interval. To test for significance the Median Test appeared to be the measure of choice for our purposes because Bartlett's Test[1] showed heterogeneity of variance was present; also raw score distributions were skewed, shifting from a marked positive skew at early age levels to a marked negative skew at upper age levels where the ceiling of the test was reached.

Each median with that of the next higher age interval was tested for significance[6]. All tests were significant ($p < .01$) except the comparisons of the two lowest age groups and the two highest age groups. The results indicate, therefore, that the total raw score discriminates adjacent six-month age interval groups between the ages of 2-6 and 5-6.

When the age interval groups were subdivided according to sex, it was noted that females tended to score slightly higher than males at each age range, with one exception. (See Table D of the Appendix.) However, the correlations between sex and median score were not significant except in the highest age group, 5-6 to 6-0, where the Phi coefficient reached .36[4].

Age interval groups were also subdivided into two categories according to parents' occupation, in order to compare the effects of socio-economic status. One group included the occupations listed under professional, managerial, clerical, and sales categories and the other group combined the occupations listed under craftsmen, operatives, service, and labor. There was a slight tendency for the children of higher socio-economic status to obtain higher scores, but the correlation between parents' occupation and CPIT score was significant only for one of the eight age-interval groups. (Age range 2-6 to 3-0, Phi coefficient .28.) It appeared, from these results, that the CPIT was, on the whole, not significantly related either to parents' occupation or sex of subject within the age range studied.

CPIT and Revised Stanford Binet. The CPIT scores of fifty subjects were correlated with their mental age ratings on the Revised Stanford Binet, Form L. A scatter plot justified the use of the product-moment correlation which was .89 $\pm$ .03. The CPIT scores of 59 handicapped children were similarly correlated with their mental age ratings. The obtained correlation was .80 $\pm$ .05.

Item Analysis and Weighting. The preliminary research had found that assigning weighted scores to the CPIT items improved the correlation with Stanford Binet scores for the group of handicapped children under investigation. There had been, at that time, no opportunity for cross-validating these weights in a replication study. The present research furnished data whereby the stability of the original weights could be tested for the two validity groups (59 handicapped, 50 non-handicapped). In neither group did the weighted scores alter the correlation between CPIT and mental age.

Several other weighting schemes were explored. One was based on the percentage passing each item in each of the eight age groups of the 400 standardization subjects. In another approach, using the same data, Phi coefficients were calculated correlating the passing of each item with total CPIT score for each of the age interval groups. On the basis of the data from the non-handicapped validity group the passing of each item was correlated with Binet mental age. (See Tables E and F of the Appendix for these data.) A number of different weighting systems were constructed using the above information in various combinations. None of them significantly altered the correlations of .89 and .80 obtained between unweighted CPIT scores and Binet mental age.

Reliability. The scores of the fifty subjects to whom two CPITs were administered were arranged for test-retest correlation. The product-moment correlation of the two sets of scores was .93 $\pm$ .02.

Since the test-retest reliability coefficient might be considered inflated due to the extended age range used, the test was split into odd- and even-numbered items; split-half reliability coefficients were calculated for one-year age interval groups, four

groups, 100 subjects in each. The four product-moment reliability coefficients, corrected according to the Spearman-Brown formula, ranged from .85 to .90.

Tables of Norms. The skewed nature of the distributions of raw scores for each age group made normalizing mandatory for most efficient test use in practical clinical situations. Normalized standard scores (T scores) were therefore calculated. Table 2 lists the CPIT raw scores and T score equivalents for the 8 six-month age interval

TABLE 2. CPIT RAW SCORES AND T SCORE EQUIVALENTS FOR SIX-MONTH AGE INTERVAL GROUPS

	Age							
CPIT Raw score	2-0 to 2-6 T	2-6 to 3-0 T	3-0 to 3-6 T	3-6 to 4-0 T	4-0 to 4-6 T	4-6 to 5-0 T	5-0 to 5-6 T	5-6 to 6-0 T
0	36	27						
1	41	34						
2	42	37		29				
3	42	38		30				
4	43	39		31				
5	46	41	29	32				
6	48	42	33	33				
7	50	44	35	34				
8	52	47	36	36		31		
9	54	48	37	37	29	32		
10	55	51	39	38	30	33		
11	57	52	41	39	31	34		
12	61	54	43	40	31	34		
13	66	57	45	41	32	36		
14	66	60	46	42	32	36		
15	67	61	48	42	33	37		
16	67	63	49	43	33	37		
17	67	63	49	43	34	38		
18	68	64	51	44	34	40	29	
19	69	65	52	45	34	40	31	
20	70	66	54	47	37	40	32	
21	72	66	55	48	38	40	32	
22	75	67	56	49	39	40	33	
23		68	57	50	40	41	34	
24		70	58	52	42	41	34	
25			59	53	43	41	34	
26			59	54	44	41	35	
27			60	55	46	42	35	
28			60	55	47	43	35	
29			61	56	49	43	36	29
30			63	57	50	43	36	33
31			65	58	52	44	37	34
32			67	59	53	44	38	34
33			69	60	54	46	40	35
34				61	56	49	42	37
35				64	59	52	44	40
36				66	60	56	46	44
37				68	63	59	48	47
38				70	67	64	51	51
39					70	70	60	58
40							74	69

groups. Table 3 lists T score equivalents for the raw scores of one-year age interval groups. In each of these T score tables a score of 50 represents the mean and the standard deviation is 10. Table 4 permits the translation of T scores into centile scores, which may be more familiar to some professional workers.

Interpretation of the results with respect to a child older than 6 years whose test score falls within the ranges obtained here is less precise. Presumably the test is appropriate for children over the age of six who are sufficiently handicapped that they

TABLE 3. CPIT RAW SCORES AND T SCORE EQUIVALENTS FOR ONE-YEAR AGE INTERVAL GROUPS

CPIT Raw score	Age 2-0 to 3-0 T	3-0 to 4-0 T	4-0 to 5-0 T	5-0 to 6-0 T
0	34			
1	37			
2	39	27		
3	40	28		
4	41	29		
5	43	29		
6	45	31		
7	47	34		
8	49	36	28	
9	51	37	31	
10	53	38	32	
11	54	40	32	
12	57	42	33	
13	60	43	34	
14	62	44	34	
15	63	45	34	
16	64	46	35	
17	64	46	36	
18	65	47	37	27
19	67	48	38	28
20	67	50	38	29
21	68	51	39	29
22	69	52	39	30
23	70	53	40	30
24	74	55	41	31
25		56	42	31
26		56	43	32
27		57	44	32
28		57	45	32
29		58	46	33
30		59	47	34
31		61	48	35
32		62	48	36
33		63	50	37
34		64	52	40
35		68	55	42
36		74	58	45
37		75	61	47
38			65	52
39			70	59
40			75	70

cannot participate in a more formal psychometric examination, and who are mentally retarded to some degree. In the absence of direct mental age equivalents for CPIT scores, at this stage of test development, the data in Table 5 permit one to compare the score of a child of any age with the obtained age-development patterns of the standardization group.

DISCUSSION

The data reported in this study establish the fact that the ability to make correct responses to the CPIT items increases with age between 2 and 6, and that the test is most effective between 2½ and 5½ years of age. Exactly what skills contribute to the ability to make correct responses cannot be defined. As the name of the test implies, it appears to measure the amount of information the child has assimilated from his environment. It has been demonstrated that the ability to apply this information to selecting the correct response is closely related to intelligence as measured by the Revised Stanford Binet. Undoubtedly correct response involves the capacity to select critically, and to inhibit immature naming responses, that is, to

TABLE 4. FOT TRANSLATING T SCORES INTO CENTILES

T score	Centile	T score	Centile
27	1	52	58
28	1	53	62
29	2	54	66
30	2	55	69
31	3	56	73
32	4	57	76
33	4	58	79
34	5	59	82
35	7	60	84
36	8	61	86
37	10	62	88
38	12	63	90
39	14	64	92
40	16	65	93
41	18	66	95
42	21	67	96
43	24	68	96
44	27	69	97
45	31	70	98
46	34	71	98
47	38	72	99
48	42	73	99
49	46	74	99
50	50	75	99
51	54		

TABLE 5. PERCENTAGE OF EACH AGE GROUP WHO EXCEED CERTAIN CPIT TOTAL SCORES

CPIT Raw score	Age 2-3	3-4	4-5	5-6
0	95%	100%	100%	100%
5	76%	98%	100%	100%
10	38%	88%	96%	100%
15	10%	69%	95%	100%
20	4%	50%	88%	98%
25		27%	79%	97%
30		18%	62%	95%
35		4%	31%	79%
39			1%	2%

hold the test demands in mind, as well as possessing the required factual information; attention and concentration also play a role. In the light of these considerations it is safe to assume that whatever the CPIT measures it corresponds with one or more facets of intelligence. CPIT score may, then, be a way of estimating intelligence in much the same way that vocabulary score is often accepted as the best single prediction of total score on the more complex psychometric instruments. Under no circumstances would this kind of measure be preferred if more extensive and varied examination is possible and available. In the setting in which the instrument was devised, for use with severely handicapped young children whose capacities could not be measured effectively by conventional means, the CPIT appeared to provide valid information which might not otherwise be available. It is anticipated, also, that the test may prove to be useful more generally by supplying confirmatory evidence in the evaluation of many preschool aged children. Since this research was initiated, Haeusserman's use of the structured interview in assessing the developmental potential of young handicapped children has been published[3]. Her very painstaking approach can provide a kind of detailed information not inherent in the CPIT. The two methods may well complement each other in clinical use according to the practical demands of a given situation.

It appears from the data obtained in this standardization study that the accumulation of a larger cross-validation population having both CPIT and conventional mental age scores might lead to the establishment of mental age equivalents for CPIT scores. There are, however, certain variables which would have to be fully investigated. It is possible that such equivalence might prove to be different for different age groups. It is also possible that equivalence between CPIT and mental age ratings might prove to be different for handicapped and non-handicapped groups, and might even be different for different kinds of handicapped populations, for example mentally retarded children and children with motor disabilities. Finally, as our philosophy of assessment and evaluation in this field continues to grow and develop, we may find that we are willing to dispense with the concept of a mental age score as making little real contribution to our understanding.

Summary

The CPIT was administered to 400 children equally divided into eight six-month age interval groups between the ages of 2 and 6. The groups were selected so that they approached as closely as possible an optimal distribution of sex, age in months, and parents' occupation. Both mean and median total scores increased from each age group to the adjacent higher age interval. The discrimination between adjacent age interval groups on the basis of median score was statistically significant between the ages of 2½ and 5½. On the whole CPIT was not significantly related either to parents' occupation or sex within the age range studied.

For fifty subjects selected from the standardization population the product moment correlation between CPIT and Revised Stanford Binet was .89; for 59 handicapped children from a variety of sources the correlation between CPIT and Revised Stanford Binet was .80.

Test-retest reliability for a group of fifty subjects was .93; split-half reliability for one year age interval groups ranged from .85 to .90.

Normalized standard scores were calculated for both six month and one year age interval groups.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1950.
3. Haeusserman, Else. *Development potential of preschool children.* New York: Grune and Stratton, 1958.
4. Jurgenson, C. E. Table for determining Phi coefficients. *Psychometrika*, 1947, *12*, 17-29.
5. Kogan, Kate L. A method of assessing capacity in preschool cerebral palsied children. *J. clin. Psychol.*, 1957, *13*, 54-56.
6. Siegel, S. *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THE PAIN APPERCEPTION TEST: AN APPLICATION TO SEX DIFFERENCES

DONALD V. PETROVICH[1]

Veterans Administration Hospital, Jefferson Barracks, Missouri

Problem

The experimental literature reporting sex differences for pain thresholds and experiences is concerned primarily with traditional psychophysical measurements. While there is tacit recognition that apperceptive factors are important in the experience of pain, little has been done to bring in this potentially rich background for more comprehensive studies. In order to initiate assessments of this apperceptive background, the Pain Apperception Test was developed [6]. Thus far, its application has been directed mainly to demonstrating its efficacy as a research tool through assessing its relationship to some indicators of anxiety and neuroticism, and by examining the elements of anticipatory pain responses [7].

Experimental literature on sex differences and pain is meager and inconclusive. Thus Chapman and Jones [1], using heat radiation with 200 normal subjects, found that the average male had slightly higher pain perception and pain reaction thresholds than the average female, with the differences not being significant. Kennard [4], using electrical stimulation, found the pain perception threshold higher in males in both patient and control groups. Hall and Stride [3] observed higher heat-pain reaction thresholds in males among psychiatric hospital admissions. On the other hand, Stengel, Oldham, and Ehrenberg [8], using physical pain stimuli such as pricking, pressure, and probing, reported "a small tendency for men to react more than women."

While these studies generally have been of the traditional psychophysical variety, it is proposed that eventually more complete studies should include not only the psychophysical measurements, but also an assessment of the predispositions that tend to determine—perhaps within wide limits—variations of these thresholds. The purpose of the study reported here was to enter the obvious but inadequately studied area of sex differences in the predispositions to experience pain.

Method

Subjects were 100 professional and clerical personnel (50 male and 50 female) of Jefferson Barracks Veterans Administration Hospital. Mean age and education, respectively, were 39.14 years and 15.32 grades for the men; and 40.18 years and 13.36 grades for the women.

The Pain Apperception Test (PAT) was used to assess attitude toward pain as reflected in judgments by each subject of the intensity and duration (each on a 7-point, multiple-choice scale) of the following 17 pictures: 1, bumped head; 2, cut forearm; 3, hit shin; 4, stomach ache; 5, smashed thumb; 6, electric shock; 7, headache; 8, stubbed toe; 9, broken leg; 10, fall on back off ladder; 11, ice dropped on foot; 12, deltoid hypodermic; 13, tooth drilling; 14, shoulder gunshot; 15, burned hand; 16, fish hook in neck; and 17, removing foreign substance from eye. Each picture was responded to for the questions "How does the man feel?" (intensity dimension) and "How long will it hurt him?" (duration dimension). The eight counterpart pictures of the PAT (four Anticipation and four Other-Inflicted) were not used in this study. Additional information regarding rationale, development, and use of the Pain Apperception Test is available in previous publications [6, 7]. Subjects were individually given the PAT, usually at their work station, with time per subject averaging approximately 15 minutes. Analysis of results was done according to statistical techniques as advocated by Edwards [2] and by McNemar [5].

[1]The writer is indebted to Dr. Rex. M. Collier, Chief Psychology Service, for active assistance in the literary presentation; and to those personnel of Jefferson Barracks Veterans Administration Hospital who acted as subjects.

Results

Split-half reliability of PAT scores was .60 for Intensity and .89 for Duration for the 50 males; and .70 for Intensity and .84 for Duration for the 50 females. Mean difference between sexes (N = 50 each) was significant for the education variable; age difference was not. There were comparatively small and non-significant mean differences between sexes in both age and education for the professional and for the clerical worker sub-groups. Results of two-tailed *t*-tests comparing the sexes as to mean PAT scores, for all subjects and for the component professional and clerical sub-groups, are presented in Table 1.

Table 1. Pain Apperception Test Scores: Seventeen Picture Totals, Comparison of Sexes

Subjects	Male			Female			
	N	M	S.D.	N	M	S.D.	*t*
Intensity							
All Ss	50	66.72	7.83	50	69.36	9.01	1.53
Professional	28	66.79	8.70	16	72.00	9.58	1.80
Clerical	22	66.64	7.38	34	68.12	9.18	.66
Duration							
All Ss	50	56.30	8.67	50	60.02	5.32	2.52*
Professional	28	55.07	9.11	16	60.88	10.75	1.82
Clerical	22	57.86	8.83	34	59.62	8.24	.75

*Significant between .05 and .01 levels.

Table 1 reveals that in all cases the females rated both the intensity and the duration of the pictured pain situations as higher than did the males. The difference was significant (P between .05 and .01) for Duration when entire groups were compared, and there was a marked tendency for the females to "see" more pain Intensity in the pictures than the males. Sex differences between clerical worker sub-groups were small whereas differences between the two professional worker sub-groups approach statistical significance (*p* between .10 and .05) for both Intensity and Duration.

Comparison of sexes as to mean scores on individual pictures revealed that the females rated 10 of the 17 pictures as signifying greater Intensity, and 14 of the 17 pictures as signifying greater Duration, than did the males. Significant differences (critical ratio) between the 50 males and 50 females in judging specific PAT pictures are shown in Table 2.

Table 2. Significant Sex Differences In Evaluation Of Pain Apperception Test Pictures (N = 50 Males, 50 Females)

Pict. No.	Pain Dimension	Male		Female		
		M	S.D.	M	S.D.	C.R.
3	Intensity	2.82	1.13	3.42	1.04	2.78**
6	Duration	2.06	.78	2.56	1.13	2.56*
10	Intensity	3.88	1.73	4.76	1.99	2.36*
10	Duration	4.66	1.35	5.48	1.43	2.95**
11	Intensity	4.98	.86	5.38	1.04	2.10*

*Significant at .05 level.
**Significant at .01 level.

The five significant differences concern four pictures (No. 3, shin hit by baseball; 6, electric shock; 10, fall on back from ladder; 11, ice dropped on foot) of varying degrees of intensity and/or duration as indicated by the mean scores. In each instance the females scored higher.

Discussion

Because of disproportionate representation of the sexes in certain jobs, and the added factors of time and availability, the subjects were fairly heterogeneous as to occupational classification. The sub-classification into professional and clerical workers was undertaken better to control the possible importance of occupational identity in regard to pain attitudes. The resulting strong tendency among professionals for females to judge the pictures as more painful than did the males, in conjunction with the negligible sex differences between clerical worker sub-groups, presents the possibility of significant occupational biases. The professional category included psychologists, physicians, social workers, rehabilitation therapists, nurses, and others, and these were not equally represented as to sex.

Evaluation of results is also complicated by the fact that all PAT pictures depict males. The importance of this in influencing and possibly producing sex differences in empathy and projection is problematical. However, the significantly greater mean Duration total response to all pictures by the entire female group (Table 1), along with their significantly greater Intensity and/or Duration response to a specific 4 of the 17 pictures (Table 2), constitutes some evidence that real sex differences in pain apperception and attitudes may exist.

Results of this exploratory study do suggest a measure of agreement with the psychophysical studies cited above, wherein higher pain perception and/or pain reaction thresholds were generally reported for the males. In the present experiment, the occupational heterogeneity of subjects precluded more definitive statistical analysis, and further experimentation along these lines should better control for this and other relevant variables. Present results also indicate that the assessment of occupational biases in pain attitudes may well prove interesting and fruitful.

Summary and Conclusions

This study was designed to initiate a new method of assessing sex differences in pain reactions. As contrasted with the traditional psychophysical method, the recently developed Pain Apperception Test was used to assess sex differences in predispositions to experience pain. Intensity and Duration judgments of 17 pictures of the PAT were obtained from 50 male and 50 female hospital personnel. Split-half reliability of this essentially projective test was acceptably high, ranging from .60 to .89. Obtained results indicated that females were generally predisposed to see more pain—both Intensity and Duration—in the PAT pictures than were the males; with this difference being significant for total Duration scores, and for a specific 4 of the 17 pictures. Similar sex differences were present but less apparent (strong tendencies) when professional worker sub-groups were compared, and negligible differences were found in comparing clerical workers. The depiction of males in all the PAT pictures presented a complicating factor, and the possible influence and importance of occupational biases in pain attitudes was noted.

References

1. Chapman, W. P. & Jones, C. M. Variations in cutaneous and visceral pain sensitivity in normal subjects. *J. clin. Invest.*, 1944, *23*, 81-91.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Hall, K. R. L. & Stride, E. The varying response to pain in psychiatric disorders: a study in abnormal psychology. *Brit. J. med. Psychol.*, 1954, *27*, 48-60.
4. Kennard, M. A. Responses to painful stimuli of patients with severe chronic painful conditions. *J. clin. Invest.*, 1952, *31*, 245-252.
5. McNemar, Q. *Psychological Statistics.* (2nd Ed.) New York: John Wiley & Sons, 1955.
6. Petrovich, D. V. The Pain Apperception Test: a preliminary report. *J. Psychol.*, 1957, *44*, 339-346.
7. Petrovich, D. V. The Pain Apperception Test: psychological correlates of pain perception. *J. clin. Psychol.*, 1958, *14*, 367-374.
8. Stengel, E., Oldham, A. J. & Ehrenberg, A. S. C. Reactions to pain in various abnormal mental states. *J. ment. Sci.*, 1955, *101*, 52-69.

AN ANALYSIS OF GROSS ERRORS ON A SENTENCE COMPLETION TEST

WILSON H. GUERTIN

Veterans Administration Hospital
Knoxville, Iowa

Problem

Sentence completion type tests provide such interesting material for content analysis that errors, *per se*, are given little attention in the final interpretation. The psychiatric patients at our hospital produce many gross errors that we would like to understand better. It was hoped that an analysis of the intercorrelations of these errors made by our patients might reveal some consistent dysfunctions that might be identifiable and of diagnostic significance.

Procedure

One hundred Rotter Incomplete Sentence Tests[1], containing at least five errors each, were selected from the files of our psychiatric hospital patients. To intercorrelate errors it was necessary for the investigator to assign them semi-objectively into categories. The errors were studied, categorized and, finally, after much trial and error, some fairly reliable categories emerged as seen in Table 1. The frequency of occurrence of a given category of errors was to be intercorrelated with each other category so cutting lines were established and tetrachoric intercorrelations were calculated. The matrix was factored by Thurstone's complete centroid method and the resultant matrix was rotated to oblique simple structure.

Results and Discussion

Table 1 shows the oblique factor matrix with a brief designation of the nature of the error-item. The four obtained factors accounted for about half of the total variance of the error-items, suggesting that errors are not just random, unsystematic occurrences but may be of diagnostic significance. Correlations between factors are all less than .20 except for −.25 between the first and second, and .43 between the first and third. While the first three factors represent tendencies to commit undesirable grammatic and syntactic errors, the fourth factor seems to represent the constructive tendency to express oneself economically and effectively.

Table 1. Oblique Factor Loadings for Error-Items

Error-item	Word Carelessness	Irrelevance	Poor Spelling	Telegraphic Phrasing
Simple word used wrongly	68	07	−01	05
Gross malapropism	57	−22	26	38
Extra word inserted	56	−23	11	−19
Simple word substitution	39	−16	17	17
Answers with phrase	−37	−14	36	−13
Disregards printed structuring	−32	88	−20	10
Repeats word perseveratively	06	48	−04	03
Total irrelevance	17	47	23	40
Final "s" omitted	21	33	31	−25
Letter substitution	−09	39	63	−15
Error in spelling	39	52	57	14
Uncrossed "t"	27	11	54	14
Omission (other)	34	48	−09	64
Tense wrong	18	−01	03	−62
Omits article	−04	51	25	52
Omits verb	45	−14	03	48
Omits conjunction	−03	13	−37	42

Table 2 presents factor loadings on some items, entirely unrelated to errors, that were introduced to clarify the nature of the obtained factors. They did not enter the actual factor analytic determinations. The first eight reference items were based exclusively upon ratings of the nature of all the sentence completion responses. The items were selected *a priori* from many possible ratings of content in the hope that they would prove most relevant to the area under investigation. The seventh and eighth items were based upon word and item counts, respectively.

TABLE 2. EXTENDED FACTOR LOADINGS FOR REFERENCE ITEMS

Reference item	Word Carelessness	Irrelevance	Poor Spelling	Telegraphic Phrasing
Delusions	18	24	20	55
Abstractions	30	−23	22	50
Complaints	62	−21	19	20
Denials	−55	34	−27	18
Perseveration of ideas	46	−10	02	05
Platitudes	−12	−25	24	−21
Productive	22	−33	−38	25
Skips items	−21	−15	−41	−20
IQ (Shipley)	−51	−32	−42	37
DQ (Shipley)	30	−15	12	−53
Education	16	−09	16	34
CA	−05	21	18	09
Paranoid schiz.	09	05	10	11
Organic B.D.	−24	03	06	−32
Chronic undiffer. schiz.	−06	09	−01	−02

Table 2 reveals the strong influence of a group of schizophrenics who are delusional, concerned with abstractions, and show a better preservation of verbal than conceptualizing ability. Since verbal ability is retained, the Shipley IQ remains higher than for those with a severe loss of verbal ability, thereby suggesting less pathology. However, the preservation of verbal ability produces a lower DQ for these patients although they actually have suffered less deterioration. These subjects employ the relatively sophisticated style of expression labeled the Telegraphic Phrasing factor.

Another cluster of correlations is encountered in Table 2 that implies the presence of individuals who complain excessively, do not deny their own problems, and have poorly preserved mental ability. They tend to perseverate with a single idea because of low ability and preoccupation with problems. These subjects are so self-focused that they fail to attend to what they write as seen by their heavy loading on the Word Carelessness factor.

SUMMARY

One hundred Sentence Completion Tests of mental hospital patients were examined for gross errors. The errors were categorized and then intercorrelated. A factor analysis demonstrated that there was consistency in the nature of the errors made and that the type of error was related to other material of clinical significance. The four factors obtained were labeled: *Word Carelessness, Irrelevance, Poor Spelling,* and *Telegraphic Phrasing.*

REFERENCES

1. ROTTER, J. B. *Incomplete Sentences* (Adult Form). New York: Psychological Corp., 1950.

DEVIANT RESPONSES AS INDICATORS OF IMMATURITY AND NEUROTICISM[1]

JOHN C. ROITZSCH AND IRWIN A. BERG

Louisiana State University

PROBLEM

Various studies have demonstrated that the distribution of subjects' responses to relatively unstructured stimuli do not follow chance distributions [1, 3, 4, 5, 6, 7, 8, 10, 11]. Typically, one finds percentages of 80-20 or some equally skewed distribution [2]. Such biased response patterns may be reflections of personality characteristics; however, attempts to use the bias or pile-up of responses as measures of personality and other traits have met with only moderate success [3, 7, 13]. But when attention is turned from the subjects who exhibit the common or popular expression of response bias to those whose responses run *counter to* the common bias, a definite relationship to personality characteristics may be discerned. Such responses which were significantly different statistically from the common response pattern have been labelled *deviant responses* by Berg [2, 3].

An examination of deviant responses as they occurred in various groups of subjects led to the formulation of the Deviation Hypothesis which has been stated as follows: "Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality (atypicalness) and thus regarded as symptoms (earmarks or signs) are associated with other deviant response patterns which are in noncritical areas of behavior and which are not regarded as symptoms of personality aberration (nor as indicators, signs, earmarks)." [3, p. 159]

A number of studies [1, 4, 5, 9, 11] of the Deviation Hypothesis have used the *Perceptual Reaction Test* (PRT) [6] as a convenient instrument for eliciting deviant responses. This test consists of 60 abstract designs drawn with ruler and compass, each with four options: *like much* (LM), *like slightly* (LS), *dislike slightly* (DS), and *dislike much* (DM), of which the subject is required to choose one. About seven minutes are required for normal subjects to complete the test.

Employing the PRT in the framework of the Deviation Hypothesis, Barnes [1] used the frequency of deviant responses to develop reliable scales for measuring general aberrant behavior, psychotic states, character disorders, and schizophrenic reactions. Hesterly and Berg [9] investigated immaturity of behavior in schizophrenia using the PRT to determine whether the deviant response patterns of adult schizophrenics would be similar to those of normal children. Their results indicated that the response patterns of groups of normal, young children (ages 7, 9, 11) to the PRT significantly differed from those of normal adults but were similar to those of adult schizophrenics. They also reported that as normal children mature, their response patterns gradually approach the patterns for normal adult responses on the PRT.

The question then arises, since the response patterns of groups of normal, young children (ages 7, 9, 11) more nearly approximate those of schizophrenic adults than that of normal adults, whether the patterns of groups of older children (ages 12-17) approximate those of neurotic adults, who are usually characterized by somewhat more mature behavior than schizophrenic adults? Therefore, according to the Deviation Hypothesis, we should expect that immature and neurotic reactions, which represent deviant responses in critical areas of behavior, would be associated with deviant responses in noncritical areas of behavior, such as atypical option preferences on the PRT. On this basis it may be initially hypothesized that a significant number of the PRT responses made by a group of normal adolescents, when compared to a group of normal adults, will be deviant and will serve to differentiate the normal adolescents from the normal adults. Similarly, as a second hypothesis, groups of adult neurotics will be distinguishable from groups of normal adults on the basis of PRT deviant response frequency. However, since normal adolescents and neurotic

[1]An expanded version of this study is available from Irwin A. Berg in mimeographed form.

adults are regarded as behaviorally immature, a third hypothesis would state that the adolescents and the neurotic adults should reveal a similar number of deviant responses. Finally, it is hypothesized that normal adults and neurotic adults will be significantly different statistically from adult schizophrenics in terms of deviant response occurrence as measured by the Barnes[1] schizophrenia scale.

Method

Three hundred white school children (150 males and 150 females) and a group of 45 white neurotic adults (20 males, mean age 37.5 and 25 females, mean age 41.2) were used as subjects for the present study. The school children were divided into three age groups, 12-13; 14-15; and 16-17, with 50 males and 50 females in each group. These adolescent subjects were obtained from various public schools in Louisiana, and all were normal in the sense of exhibiting no obvious signs of psychopathology. The adult neurotic subjects were drawn from psychiatric clinics in Louisiana and Texas and had received a staff diagnostic label of "psychoneurosis". The normal adult subjects were 500 white males and 350 white females drawn by Barnes[1] from a norm group of more than 4,000 subjects in Berg's files. The PRT data for schizophrenic patients were also obtained from Barnes's study.

The PRT was administered to the children in their classroom and to the neurotic adults in the psychiatric clinics. The PRT responses for each subject were transferred to IBM cards for scoring and tabulation. For every PRT item, the four option choices made by each of the three groups of adolescents were separately summed for male and female patients. The same procedure was followed for the adult neurotic and normal groups. These frequencies of option choices were then converted into percentages for each group with males and females tabulated separately. The percentage of normal adult subjects choosing each option was then compared with the percentage of adolescent subjects in each of the three groups who chose the same option. The same comparison was made for the adult normal with adult neurotic groups and for the adult neurotics with each of the adolescent groups. Each comparison required testing the level of statistical significance of the difference in the percentages for each option by use of chi-squares in a contingency table as described by Mainland and Murray[12]. Since the PRT has 240 options and 14 comparisons between the various groups were required, the level of statistical significance for 3,360 options was determined in the course of the present study. Finally, to determine whether the neurotic groups were different from schizophrenics, the PRT choices for normal and neurotic groups were scored on the Barnes[1] PRT schizophrenia scales (called *Sigma Scales* by Barnes) and the significance level of the mean schizophrenia scores for each group was determined.

Results and Discussion

On the basis of chance alone, two or three of the 240 PRT options would occur as deviant responses when the one per cent level of statistical significance is used in comparing groups. It was found that this chance expectancy was vastly exceeded. From 74 to 144 of the 240 PRT options differentiated groups of normal adolescents from normal adults at the one per cent level of significance.

When neurotic adults are compared to normal adults 132 options for the neurotic males and 120 options for the neurotic females were significantly different at the one per cent level when compared to the normal adult males and females. The data in Table 1 summarize the results of scoring the normal, neurotic, and schizophrenic groups with Barnes' PRT schizophrenia scale. It is evident that both male and female neurotics are significantly different from male and female schizophrenics. The neurotic females are also significantly different from normal females in schizophrenia score. The neurotic males, however, do not differ in this respect when compared to normal males.

Thus, as hypothesized, normal adolescents and adult neurotics both differ significantly from normal adults when compared on the basis of deviant responses to

TABLE 1. MEAN PRT SCHIZOPHRENIA SCORES FOR GROUPS OF NORMAL, PSYCHONEUROTIC, AND SCHIZOPHRENIC ADULTS

	Normal[a]		Neurotic		Schizophrenic[a]	
	Males (N = 500)	Females (N = 350)	Males (N = 20)	Females (N = 25)	Males (N = 99)	Females (N = 68)
Mean Schizophrenia Score	−19.43[b]	−15.03[b]	−14.25[b]	−3.56[b,c]	2.52[c]	6.14[c]

[a]The data for normal adults and schizophrenic adults are from Barnes[1].
[b]Significantly different from the mean score of the schizophrenic group at the one per cent level.
[c]Significantly different from the mean score of the normal group at the one per cent level.

the PRT. The normal adolescents and neurotic adults, however, have a quite similar deviant response frequency. Hence, adult neurotics are immature in the sense that their PRT responses are like those of 12 to 17 year-old children rather than like those of normal adults. The deviant response pattern of neurotic adults does not resemble that of schizophrenics, for the two groups are readily distinguished from each other by the Barnes schizophreniza scale.

SUMMARY

The Deviation Hypothesis holds that deviant responses are general; hence deviant responses in a critical area of behavior (*i.e.*, neurotic symptoms) will be associated with other deviant responses in noncritical areas of behavior (*i.e.*, choice of test options). As a test of one facet of this hypothesis, the Perceptual Reaction Test (PRT) was administered to 300 normal adolescents and 45 neurotic adults. Statistically significant differences in the frequency of deviant responses were found between groups of normal adolescents and normal adults. Significant differences were not found between groups of normal adolescents and neurotic adults, indicating that neurotic adults reveal PRT deviant response frequencies which are similar to those of normal adolescents. Thus, neurotic adults exhibit immature response patterns in the sense that they are quite similar to those of normal adolescents as measured by deviant response frequencies on the PRT. The deviant response pattern of neurotic adults was found to be significantly different from that of adult schizophrenic patients. The evidence from the present study indicates that groups of normal adolescents and groups of neurotic adults can both be differentiated from normal adults on the basis of deviant responses occurring on the PRT. Also, groups of normal and neurotic adults can be distinguished from adult schizophrenics on the basis of deviant response patterns.

REFERENCES

1. BARNES, E. H. The relationship of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, *51*, 286-290.
2. BERG, I. A. Response bias and personality: The Deviation Hypothesis. *J. Psychol.*, 1955, *40*, 61-71.
3. BERG, I. A. Deviant responses and deviant people: the Formulation of the Deviation Hypothesis. *J. counsel. Psychol.*, 1957, 4, 159.
4. BERG, I. A. The reliability of extreme position response sets in two tests. *J. Psychol.*, 1953, *36*, 3-9.
5. BERG, I. A. and COLLIER, J. S. Personality and group differences in extreme response sets. *Educ. psychol. Measmt.*, 1953, *13*, 164-169.
6. BERG, I. A., HUNT, W. A., and BARNES, E. H. *The perceptual reaction test.* Evanston, Ill.: 1949.
7. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, *6*, 475-494.
8. FRENCH, J. E. Children's preferences for abstract designs of varied structural organization. *Elem. Sch. J.*, 1956, *56*, 202-209.
9. HESTERLY, S. O., and BERG, I. A. Deviant responses as indicators of immaturity and schizophrenia. *J. consult. Psychol.*, 1958, *22*, 389-393.
10. KATZELL, R. A. Cross-validation of item analyses. *Educ. psychol. Measmt.*, 1951, *11*, 16-22.
11. LEWIS, N. A., and TAYLOR, J. A. Anxiety and extreme response preferences. *Educ. psychol. Measmt.*, 1955, *15*, 111-116.
12. MAINLAND, D., and MURRAY, I. M. Tables for use in fourfold contingency test. *Science*, 1952, *116*, 591-594.
13. RUNDQUIST, E. A. Response sets: a note on consistency in taking extreme positions. *Educ. psychol. Measmt.*, 1950, *10*, 97-99.

AN INTERNAL CONSISTENCY ANALYSIS OF THE FORCED-CHOICE ANXIETY SCALE

JOHN A. HAMMES AND E. DELL YOUNG[1]

University of Georgia

PROBLEM

Heineman [2] has revised the true-false Taylor Anxiety Scale to a multiple-choice format that minimizes the response bias of true-false items. Hammock [1] subsequently administered a revision of the Heineman scale to Army trainees, in the interest of relating A-scores to personnel selection. The resulting normative data were found to approximate Heineman's findings to the degree that Hammock concluded "it is unlikely that the Army and college groups differ on manifest anxiety." The present study consists of an internal consistency analysis of Hammock's revision of the Heineman scale, and a comparison of the present sample with those of Heineman and Hammock.

METHOD

Hammock's scale was administered as an "attitude check list" to 302 students taking introductory psychology at the University of Georgia, and the upper and lower 27 per cent of the distribution were selected. For each scale item, there were computed the percentages of these two groups checking the item. Item validity coefficients were then obtained by entering these percentages in Flanagan's table [4], which gives estimates of the product-moment correlation between the two underlying normally distributed variables. An item analysis against extreme groups was carried out [3].

The A-scale was administered with the following instructions: "In this questionnaire you will read statements describing various attitudes and feelings. The statements are arranged in blocks of three. Read the three statements and pick out the one that describes your attitude or feeling *most* and mark it here (illustrate with sample items). Then, from the remaining two statements, choose the one that describes your attitude *least* and mark it here (indicate). Then go on to the next block of three statements, and repeat the procedure. Remember, there are no right or wrong answers. Your score is your own, because everyone is different. Be sure to mark all blocks. Any questions?"

RESULTS

Item validity coefficients for the 50 anxiety items are presented in Table 1, in an arbitrary grouping. There was one item which correlated negatively, namely,

TABLE 1. INTERNAL ITEM VALIDITY COEFFICIENTS FOR THE FIFTY ITEMS OF THE FORCED-CHOICE A-SCALE

Range of Coefficient	Frequency
.75 – 1.00	0
.50 – .74	18
.25 – .49	25
.00 – .24	6
−.24 – .00	1*

*−.12

−.12. The next higher item correlated .11, and the highest correlated .63. Forty-three of the items were above .25. Items correlating below .25 will be revised for improvement.

[1]Now at Human Research Unit 3, Fort Benning, Georgia.

Surprisingly, the score distribution of the present sample differed significantly from those of Heineman and Hammock (Table 2; Figure 1). Reliability coefficients of the three samples are comparable (Table 2).

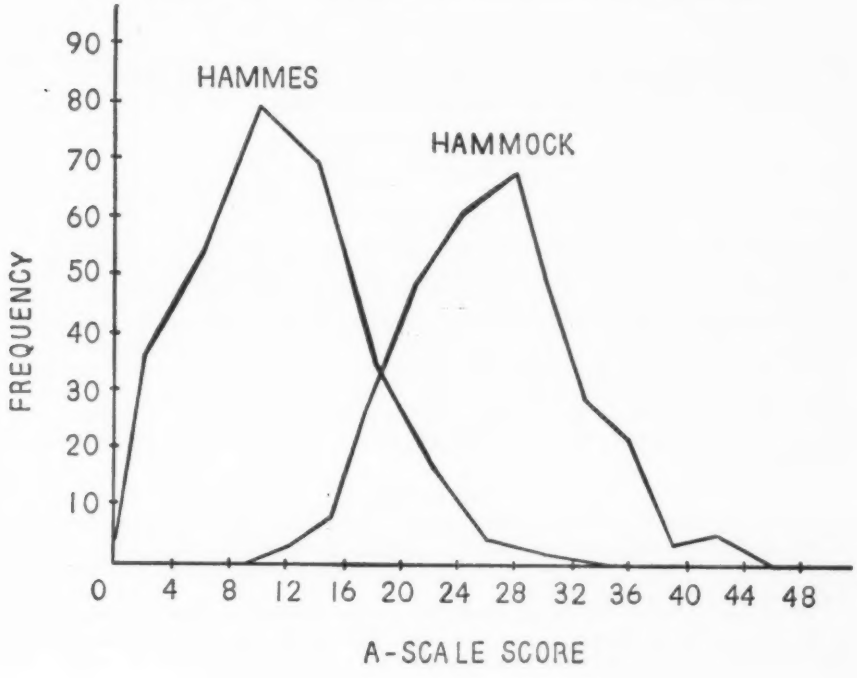
TABLE 2. MEANS AND RELIABILITY COEFFICIENTS OF THE FORCED-CHOICE A-SCALE FOR THREE SAMPLES

Sample	N	Mean	P	S.D.	P	*r*	S.E.m
Heineman	209	25.58		6.22		.69	.04*
			<.001		<.001		
Hammock	315	26.18		5.59		.58	.04**
			<.001		<.001		
Hammes	302	11.56		5.82		.62	.04**

*Kuder Richardson Formula 21.
**Odd - even split half.

It is unlikely that Heineman's sample and the present sample differ so greatly in manifest anxiety, since both were drawn from introductory classes in psychology. Since the present sample achieved lower scores on Hammock's A-Scale than did Army trainees, it might be presumed that Heineman's sample would have also. Have college students, then, less manifest anxiety than Army trainees? There is another possible explanation for score differences, in terms of Hammock's revision procedures. The forced-choice blocks of items in Hammock's scale were constructed on the basis of item favorability ratings made by Army cadre, whereas Heineman used ratings by college students. Perhaps college students are more alert to favorability bias than are Army personnel and consequently achieve lower scores on a scale prepared on the

FIG. 1. FREQUENCY POLYGONS OF FORCED-CHOICE A-SCORES FOR TWO SAMPLES.



basis of ratings by Army personnel than they would on one involving ratings by college students. The two scales would consequently not be comparable in *score* value. In summary, the difference between scores of the Hammock sample and the present sample may be due to (a) differences in manifest anxiety, or (b) differences in scale construction, or (c) possibly both factors. The question, then, of equivalence of manifest anxiety in college and Army samples remains open.

Summary and Conclusions

An internal consistency analysis was carried out on the Hammock version of the Heineman Forced-Choice Anxiety Scale, using the upper and lower 27 percent of a sample consisting of 302 college students. It was found that the forced-choice form of the true-false anxiety scale has relatively high internal consistency, and is therefore a desirable format for such a questionnaire. It was also found that the equivalence of manifest anxiety in college and Army samples is questionnable.

References

1. Hammock, J. C. Anxiety scales for use in Army research. Human Resources Research Office, Washington, D. C., 1954.
2. Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, *17*, 447-455.
3. Kelley, T. L. The selection of upper and lower groups for the validation of test items. *J. Educ. Psychol.*, 1939, *30*, 17-24.
4. Thorndike, R. L. *Personnel selection; test and measurement techniques.* New York: John Wiley & Sons, 1949.

FORMBOARD COMPARABILITY IN THE COLORED PROGRESSIVE MATRICES

THOMAS E. JORDAN

Washington University

Problem

Several studies[2, 3, 5] have suggested that the Colored Progressive Matrices[4], a measure of intellectual ability, has considerable value. Clinical application of the scale has led to consideration of the value of the incomplete designs, which are the substance of the scale, in another context; specifically, the provision of kinaesthetic and depth cues in the format of a formboard. Such a medium would also provide a slightly more provocative basis for the subjective estimates of problem solving techniques which are a valuable adjunct to the normal, objective statement of test results. The purpose of the study was to evaluate the comparability of a formboard version of the Colored Progressive Matrices to the conventional booklet version of the scale.

Procedures

Thirty-six formboards were constructed, identical in all respects except that (a) the incomplete design was recessed to a depth of one-eighth of an inch, and (b) the possible choices to complete the design were made of wood, and all were of such a shape and size as to fit the recess.

Two groups (N = 49) of school children were derived randomly and matched to provide groups comparable by mental age. One group was given the formboard scale, the other group took the booklet version.

Results

Table 1 presents CA's, MA's, IQ's and comparable booklet and formboard raw scores. With a t-value of 1.51, the difference between the means of the booklet and

Table 1. CA's, MA's and IQ's of the Subjects (N = 49)

		Range	Mean	S. D.
CA (Months)				
Group I		99–111	101.60	3.90
Group II		97–108	102.24	4.41
MA (Months)				
Group I		98–138	116.40	11.08
Group II		98–146	120.72	11.55
IQ				
Group I		93–127	110.28	9.41
Group II		91–138	114.64	11.70
Booklet	Raw Scores	16–28	21.72	3.80
Formboard	Raw Scores	18–30	23.40	3.29

formboard scores does not approach significance. The coefficients of reliability (split half) were .698 for the booklet form and .675 for the formboard. The coefficients of validity, defined as Pearson r against Stanford-Binet MA's, were $-.033$ for the booklet form and .203 for the formboard, both quite low and strikingly at variance with results of other investigators[3, 5].

Summary

This study compared test scores on the booklet form of the Coloured Progressive Matrices with a new formboard version. The reliabilities of the two forms were comparable, and the difference between the mean raw scores of the two forms did not approach significance, so it is concluded that booklet and formboard versions are comparable. Both scales produced low correlations against Stanford-Binet scores.

References

1. Jordan, T. E. and Bennet, C. M. An Item Analysis of the Colored Progressive Matrices. *J. consult. psych.*, 1957, *21*, 222.
2. Jordan, T. E. The Utility of the Colored Progressive Matrices. Unpublished dissertation, Indiana University, 1955.
3. Martin, A. and Wiechers, J. E. Raven's Colored Progressive Matrices and the WISC. *J. consult. psych.*, 1954, *18*, 143-44.
4. Raven, J. C. *The Colored Progressive Matrices.* Los Angeles, California: Western Psychological Services, 1952.
5. Smith, L. M. The Progressive Matrices and the Stanford-Binet with Exceptional Children. *J. Except. Children* 1958, 24, 374-375.

A SHORT FORM OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

WILLIAM H. SIMPSON AND CECIL C. BRIDGES, JR.

Oklahoma City Public Schools

The particular combination of subtests used in the present study (Vocabulary and Block Design) was chosen because a sum of the scores of these two subtests showed high correlation with total weighted scores in the three age groups for which Wechsler[3] presents tables of intercorrelations, using McNemar's formula[2]. The sample in the present study consisted of 120 children referred to the Division of Child Guidance of the Oklahoma City Public Schools. This sample is fairly typical of the population of children who come to the attention of social agencies and child guidance clinics. Table 1 presents descriptive data concerning general intelligence and chronological age of the subjects in this sample.

TABLE 1. MEAN, STANDARD DEVIATION, AND RANGE FOR CHRONOLOGICAL AGE AND WISC FULL SCALE IQ FOR THE SUBJECTS

Factor	Range	Mean	S. D.
Chronological Age (mo.)	65–192	124.23	26.06
Full Scale IQ	54–142	95.10	15.87

The correlation of the Vocabulary-Block Design sum with the full weighted score was .874. The obtained relationship seems strong enough to warrant the use of a short form composed of these two tests as a screening instrument. When contrasted with results of studies of short forms of the Wechsler Bellevue Intelligence Scale, the present findings are at least as favorable as those obtained with the WBIS[1]. Table 2, calculated from the regression equation and the WISC table for converting full weighted scores to IQ's, is given for converting the sum of the weighted scores of Block Design and Vocabulary tests to an estimated full-scale IQ.

TABLE 2. CONVERSION OF SUM OF WEIGHTED SCORES ON BLOCK DESIGN AND VOCABULARY TO PREDICTED FULL SCALE IQ

Blocks & Vocabulary	Predicted Full Scale IQ	Blocks & Vocabulary	Predicted Full Scale IQ
1	51	21	104
2	54	22	107
3	56	23	109
4	59	24	112
5	62	25	115
6	64	26	117
7	67	27	120
8	70	28	123
9	72	29	125
10	75	30	128
11	77	31	131
12	80	32	133
13	83	33	136
14	85	34	138
15	88	35	141
16	91	36	144
17	93	37	146
18	96	38	149
19	99	39	152
20	101	40	154

REFERENCES

1. HERRING, F. H. An evaluation of published short forms of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1952, *16*, 119-123.
2. MCNEMAR, Q. On abbreviated Wechsler-Bellevue scales. *J. consult. Psychol.*, 1950, *14*, 79-82.
3. WECHSLER, D. *Wechsler Intelligence Scale for Children.* New York: Psychological Corporation, 1949.

BASE RATE DATA ON RORSCHACH CARD REJECTION*

WALTER F. MC KEEVER AND ALVIN I. GERSTEIN

University of Rochester

Problem and Procedure

Although Rorschach card rejection is one of the more gross and easily denotable responses to the test, such behaviors have received little systematic research attention. The studies of Mensh and Matarazzo[3] and of Sisson *et al.*[6], have provided the only substantial data on card rejection. The present study, utilizing a large group of Rorschach protocols, aims primarily at determining if there are significantly different rejection frequencies between the cards. In addition, the relationships between rejection, and reaction times and IQ respectively, will be assessed.

All cases having been administered Rorschachs at the Rochester VA Mental Hygiene Clinic were taken from the test files, and rejection frequencies were tabulated for the ten cards. The only criterion for accepting a record for study was that it have at least ten responses. Any card failing to evoke a response properly scorable in the genetic framework, as presented by Friedman[2], was counted as a rejection.

The rejection data are based on an N of 693 males of various diagnoses, this number representing the entire Clinic population as defined for study. In addition, 424 of these patients had been given either Full Scale or prorated Wechsler-Bellevue I tests, permitting assessment of the relationship between IQ and card rejection. The mean IQ of these patients was 106.03, and the mean number of Rorschach responses for the entire group was 21.8. A total of 79 records had less then 10 responses and are not included in the N of 693.

Results and Discussion

Table 1. Numbers and Percentages of Rejections of Each Rorschach Card

Cards	I	II	III	IV	V	VI	VII	VIII	IX	X	Sum
Number of rejections	2	22	4	18	8	54	43	4	57	26	238
Percentage of Rejections	.8	9.2	1.7	7.6	3.4	22.7	18.1	1.7	23.9	10.9	100

The essential base rate data are summarized in Tables 1 and 2. Accepting the 5% level as the criterion for significance, individual card-to-card chi square analyses (corrected for small cell frequencies where necessary) reveal three distinct groupings of the cards, with no significant differences within groups. Thus, cards IX, VI, and VII are *all* rejected significantly more often than cards X, II, or IV. *Each* of these six in turn is rejected significantly more often than III, VIII, or I. Card V, while not falling within any of these groups, is significantly different from all cards of greater rejection frequencies, except card IV. In the population studied, rejection behavior is not uncommon, since essentially 25% of the group reject cards. The rejection of certain cards, however, such as I, III, and VIII is so infrequent as to be considered markedly deviant. From the multiple-rejection data summarized in Table 2, it can be seen that the rejection of more then two cards is extremely rare, and no one rejects more than four cards.

*Appreciation is acknowledged to the staff of the Rochester Veteran's Administration Clinic, Rochester, New York, for their cooperation in making this study possible.

TABLE 2. NUMBERS AND PERCENTAGES OF PATIENTS REJECTING ZERO TO FOUR CARDS

	Number of Rejections					
	0	1	2	3	4	Total
Number of Patients Rejecting	522	123	34	9	5	693
Percentage of Patients Rejecting	75.3	17.7	4.9	1.3	.7	99.9

Table 3 presents findings on the relationship of IQ to rejection behaviors. The point biserial correlation between IQ and card rejection is −.12, significant at the

TABLE 3. NUMBERS, MEAN IQS, STANDARD DEVIATIONS, AND t VALUES FOR THE IQ DIFFERENCES FOR NON-REJECTORS VERSUS REJECTORS OF ONE CARD AND MULTIPLE-REJECTORS

	N	Mean IQ	SD	*t*
Non-Rejectors *vs.*	314	108.2	14.42	
Rejectors of One	77	104.5	13.41	2.35*
Multiple-Rejectors	33	103.3	16.39	1.66*

*Significant at better than the 5% level.

5% level. Non-rejectors are significantly more intelligent than either category of rejectors.

Rank order correlations between obtained rejection ranks and various reaction time and rejection ranks reported by other investigators[1, 3, 4, 5, 6, 7] are significant for all adult groups (5% level or better), and range from .75 to .97. Indeed, it appears that rejection ranks correlate as highly with reaction time ranks as rejection ranks correlate among themselves. Clearly, those cards with the longest reaction times are those most frequently rejected.

Chi square reveals no significant difference between the colored versus non-colored cards, while the difference between the 5 "broken" and 5 "unbroken" blots is significant at the 1% level. The latter finding is consistent with Friedman's[2] assumption that "unbroken" blots are perceptually more difficult to structure.

CONCLUSIONS

The results of the present study are consistent with those reported for rejection by other investigators[3, 6] studying different clinical groups. Although IQ is negatively correlated with rejection, the relationship accounts for only 1% of the variance. Conceivably, tests of more specific abilities rather than the global one used here would show a greater relationship to rejection. That there is a significant difference in rejections on the dimension of "brokenness" suggests that perceptual analyzing and integrating abilities might be two such factors. That other factors and interactions are probably involved in producing rejections is indicated since "brokenness" does not maximally differentiate the most from the least rejected cards. One might reasonably hypothesize that the observed rejection pattern is at least partially a function of an accumulation of reactive inhibition interacting with stimulus complexity, their effects being manifested among individuals within some critically low range of overall drive levels; such a "fatigue" factor seems possible in view of the fact that 77% of all rejections occur on the last five cards.

The consistently similar card rankings on rejection and reaction time suggests that these variables have little diagnostic value in and of themselves, since the correlations are similarly high regardless of the clinical groups involved. This does not, of course, preclude the possibility that there are important personality correlates of rejection patterns. That children have a quite different rejection pattern[6] suggests

the possibility of developmental sequences of such patterns, which might serve as an index of genetic maturity.

References

1. Beck, S. J., Rabin, A. L., Thiesen, W. G., Molish, H. and Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, *30*, 241-298.
2. Friedman, H. Perceptual regression in Schizophrenia. An hypothesis suggested by the use of the Rorschach test. *J. Proj. Tech.*, 1953, *17*, 171-186.
3. Mensh, I. N. and Matarazzo, J. D. Rorschach card rejection in psychodiagnosis. *J. consult. Psychol.*, 1954, *18*, 271-275.
4. Phillips, L. and Smith, J. G. *Rorschach interpretation: advanced technique.* New York: Grune and Stratton, 1953.
5. Sanderson, H. Norms for shock in the Rorschach. *J. consult. Psychol.*, 1951, *15*, 127-129.
6. Sisson, B. D., Taulbee, E. S. and Gaston, C. O. Rorschach card rejection in normal and psychiatric groups. *J. clin. Psychol.*, 1956, *12*, 85-88.
7. Uehling, H. F. Rorschach "shock" for two special populations. *J. consult. Psychol.*, 1952, *16*, 224-225.

MMPI PROFILES OF PARENTS OF CHILD PSYCHIATRIC PATIENTS[1]

LEO J. HANVIK AND MILDRED BYRUM

Washburn Memorial Clinic
Minneapolis 5, Minnesota

Problem and Method

The staff of the Washburn Memorial Clinic, a child guidance clinic, have been using the MMPI routinely (4) in the assessment of the children's parents' personalities. The present study involves all of the profiles obtained on parents of children referred to the clinic in a two year period. The total sample of profiles included 104 father profiles and 158 mother profiles (except that in the portion of the study dealing with the relationship between parent MMPI patterns and the children's syndromes, the numbers were 96 fathers and 156 mothers). The larger number of mothers than of fathers resulted partly because a number of the children had no father in the home and partly because some of the fathers did not participate in the clinic treatment program or were not as cooperative as the mothers in meeting requests of the clinic.

In the study two methods of profile analysis were used: (a) mean T scores were computed on all scales and (b) the profiles were all coded by the ranking method and the mean ranks for all scales were computed for the mothers and for the fathers. The mean profiles in Figs. 1 and 2 as well as the individual profile patterns were then inspected and interpreted. In an effort to note possible relationships between parent MMPIs and child syndromes, the child of the parents studied was in each case classified according to a problem classification table previously devised at the clinic for another study. An analysis was then carried out in an effort to determine what types of parent profile codes appeared to be associated in undue proportion in any of the child problem categories.

Results

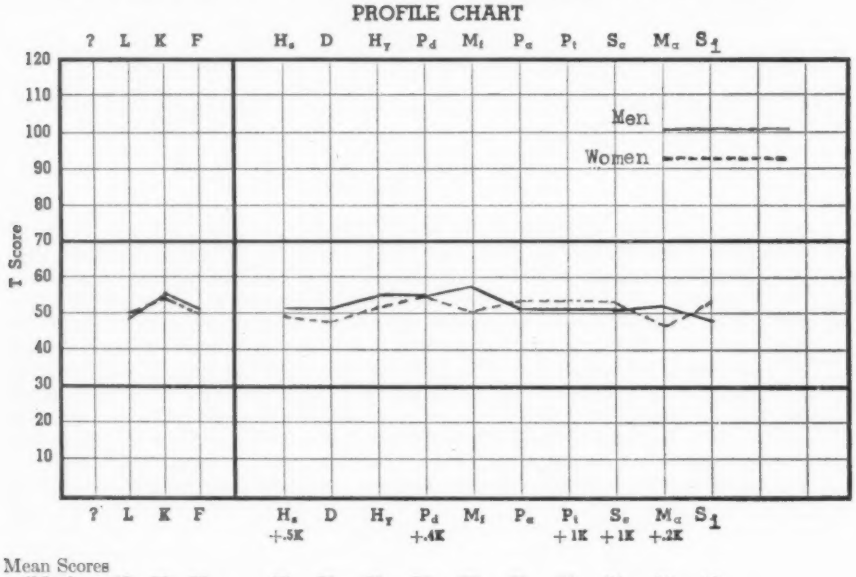
Comparison of Profiles. Composite profiles for the parents are plotted in Fig. 1 together with mean T scores on all scales for 158 mothers and 104 fathers, showing that both the mothers and fathers, as groups, tend to score above average on the Hy, Pd, Pa, Pt and Sc, with the strongest trends showing up on the Hy and Pd scales. Fig. 2 shows that in terms of relative scale relationships or of prominance of position within the individual profiles, the Hy and Pd scales occupy the first and second

[1]The authors are indebted to Mr. William H. Dressler and to other members of the social work staff of the Washburn Memorial Clinic for their work in obtaining MMPI profiles on parents as a part of the clinic's intake process. The authors also wish to express their indebtedness to Dr. Mollie MacEachern, also of the Washburn Memorial Clinic for her advice and other assistance in this study.

positions respectively in both the mothers' and fathers' profiles, omitting the Mf scale for the moment. The predominance of Hy and Pd scale elements reflected in the above mean T scores and mean scale rank patterns of the parents' MMPIs was readily observable also in an inspection of the individual profiles. For example, the "conversion V" configuration[2] was found to be present in 24% of the father's profiles and in 33% of the mothers' profiles. With regard to Pd scale prominence in individual profiles, 13% of the fathers and 18% of the mothers had Pd scale T scores over 70. Although the statistical significance of the above incidences is undeterminable, they are quite obviously high and bear out the heavy weighting of Pd and Hy scale elements in Figs. 1 and 2 illustrating mean profile T scores and ranks.

A striking feature of Figs. 1 and 2 and also discernible upon inspection of the individual profiles, was the Pd minus Ma differential which in some individuals,

FIG. 1 THE MEAN T SCORES ON MMPI SCALES FOR 158 MOTHERS and 104 FATHERS



	?	L	K	F	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	S1
Mean Scores (Men)	48	56	52		51	51	56	55	58	51	51	51	52	48
Mean Scores (Women)	50	55	50		49	47	52	55	50	53	53	52	48	52

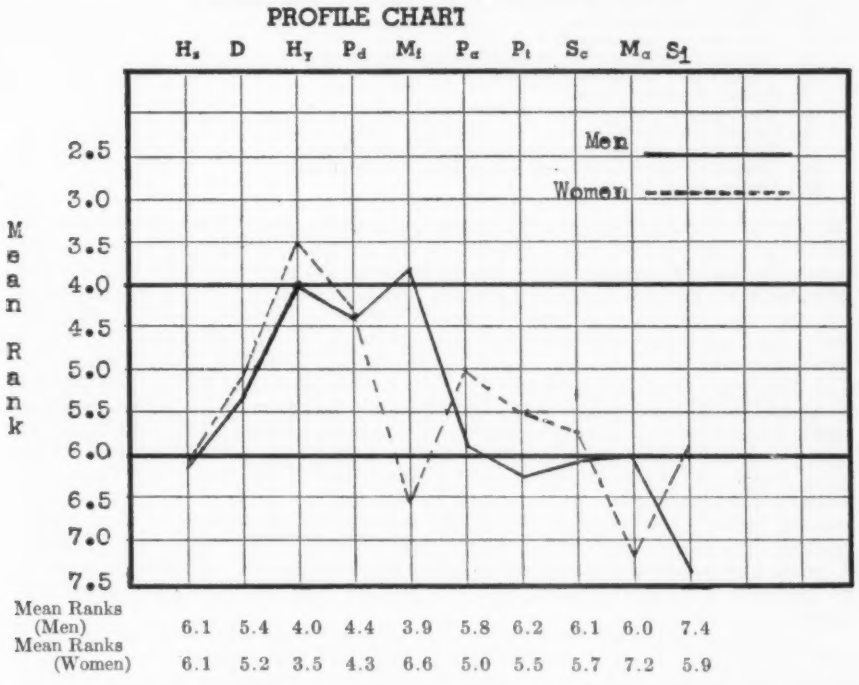
particularly mothers, was extremely high. (In one mother's profile, for example, the Pd score was 49 T score points higher than the Ma score. This mother, incidentally, showed no "acting out" trends of a generalized sort but had a severe conflict with both her husband and child.) The Pd minus Ma differential was significantly greater in the mothers than in the fathers, in both samples being significantly greater than zero.[3] The relatively high Pd score and relatively low Ma score in parents' profiles

[2] A conversion V was defined for our purposes as a profile in which the Hs scale and Hy scale both yielded T scores at least half standard deviation above the T score on the D scale. This pattern previously [2, 5] has been shown to occur in persons who use repression as a defense mechanism and who manifest anxiety in dissociated or "conversion" symptom patterns.

[3] The mean Pd minus Ma differential for the fathers group (N 103) was 3.8 with a S.D. of 11.4. The mean Pd minus Ma differential for the mothers group (N 157) was 9.7 with a S.D. of 11.3. The t for the difference between these means was 4.7 with a probability of less than .001. The fathers mean Pd minus Ma differential was significantly different from zero with a t of 3.38 which has a probability of less than .001.

has been of interest to our staff because it seemed to occur so frequently in situationally maladjusted individuals reacting adversely and with intense hostility to a portion of the social environment, *i.e.*, persons with severe marriage problems. Nelson[6] has suggested a 15 point or greater Pd-Ma differential as pathological on the basis of work with adult psychiatric patients who, though not acting out hostility in a generalized "psychopathic" fashion in the community, nevertheless are feeling and acting out hostility in a more specific fashion, usually against a family member. In our parent sample, we also frequently noted elevations on the Si and D scales in profiles with high Pd and low Ma scores, specifically in several parents presenting the most severe marital conflict problems.

FIG. 2. THE MEAN RANK PROFILES FOR 158 MOTHERS AND 104 FATHERS ILLUSTRATING THE RELATIVE PROMINENCE OF POSITION OF THE VARIOUS MMPI SCALES.



Mean Ranks (Men)	6.1	5.4	4.0	4.4	3.9	5.8	6.2	6.1	6.0	7.4
Mean Ranks (Women)	6.1	5.2	3.5	4.3	6.6	5.0	5.5	5.7	7.2	5.9

Figs. 1 and 2 present some interesting contrasts between mothers and fathers on the mean T scores and mean ranks on the Mf, Ma and Si scales, where all the mean curves show an opposite tendency of the father *vs.* the mother group.

The fathers as a group tend to have elevated Mf scores whereas the mothers tend to have depressed Mf scores. The clinic staff noted repeatedly that many of the mothers with low Mf scores (either absolutely or in relative terms within the client's own profile) were rather disturbed persons and especially difficult to influence therapeutically. To check the impression of low Mf as a disturbance index, we compared the frequency of Mf scores below 40 among the following groups of profiles: (a) those with one or more clinical scales (exclusive of the Mf scale) in the above 65 T score range, (b) those with one or more clinical scales in the 61 to 65 T score range, and (c) those wherein the clinical scales were all 60 or below. The tabulated frequencies were compared to ascertain the association of low Mf scores with higher elevation pro-

files with expected frequency by means of Chi Square which turned out to be 19.7, statistically significant and with a probability of less then .01. Our interpretation from this comparison was that mothers with the most elevated profiles tended more frequently to have Mf scores below 40 than did mothers with less elevated profiles, confirming previous impressions that low Mf scores in women are themselves signs of emotional disturbance and tend to be associated with high scoring tendencies on other MMPI scales. The fathers, on the other hand, have high Mf tendencies and there is no evidence of the fathers' Mf scores being associated with either clinical or MMPI signs of emotional disturbance.

Clinical Classifications. Each child of the parents in this study was classified according to a classification system including the following primary problem categories:

1. Rebelliousness at home or school
2. Chronic anxiety or fear reactions (excluding school phobia syndrome)
3. Habitual manipulations
4. Delinquency (adjudicated)
5. Primary learning problem or school underachievement
6. Primary peer relationship problem
7. School phobia syndrome
8. Psychosomatic complaints
9. Chronic withdrawal
10. Enuresis (as primary presenting complaint at the clinic)
11. Encopresis (as primary presenting complaint at the clinic)
12. Chronic sleep disturbance

Only rarely was the primary problem the sole complaint, but rather in most instances was the problem of paramount concern of the parent in bringing the child to the clinic. To discover possible relationships between the parent MMPI findings and their children's problem areas, the authors jointly studied the child's case record in each case and arrived at a joint diagnostic sorting or classification of each child. From this work it appears tentatively that parent MMPI profile patterns or elevations were associated with particular syndromes or presenting complaints in the child patients. There is a tendency for the mothers who have Hy as their highest scale and the Mf scale as their own first or second low point to have their children classified in category 1 in the scheme, *i.e.*, with a main problem of "rebelliousness at home or school". Chi Square is 7.008, significant at the 5% probability level before Yates correction for continuity but just missing the 5% level after the Yates correction.[4] The relationship between high Hy scores alone or low Mf scores alone as mothers' MMPI features and the child's membership in category 1 just misses significance at the 5% level even without application of Yates correction. (The low Mf score had previously been seen to occur frequently in the mothers' profiles as had been a high Hy score, hence, these two features were combined and studied as a cluster.)

In addition to the relationship between high Hy scale score and low Mf scale score and category 1 in the child problem areas, there also appeared to be a strong tendency for children with a diagnosis of "school phobia syndrome" to have mothers with MMPI codes beginning with 2, *i.e.*, with D as the high scale within the profile, although Chi Square again just misses the 5% level. No strong relationships of any sort appeared between the fathers' MMPI profiles and children's problems.

Summary

The MMPI was administered routinely to parents of children applying for treatment at a psychiatric clinic for children. Deviations from the normal appeared in the parents' profiles and there appears to be some association between certain parent profiles and certain types of presenting complaints in children. The various features of child psychiatric patients' parents' MMPI profiles are discussed in detail.

[4]This correction (2) has the function of correcting for small theoretical cell frequencies and is especially recommended when the Chi Square without correcting comes out just above customary confidence levels.

References

1. Aaronson, B. S. A comparison of Two MMPI Measures of Masculinity-Feminity. *J. clin. Psychol.*, 1959, *15*, 48-50.
2. Guilford, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill Book Co., 1950.
3. Hanvik, L. J. MMPI Profiles in Patients with Low Back Pain. *J. consult. Psychol.*, 1951, *4*, 350-353.
4. Hathaway, S. R. and McKinley, J. C. *Minnesota Multiphasic Personality Inventory: Manual.* New York: Psychological Corporation, 1951.
5. Hovey, H. B. Somatization and Other Neurotic Reactions and MMPI Profiles. *J. clin. Psychol.*, 1949, *5*, 153-157.
6. Nelson, Sherman. Unpublished study reported in a personal communication.

CHANGES IN RECOGNITION THRESHOLDS ASSOCIATED WITH CHLORPROMAZINE, PROMAZINE AND PHENOBARBITAL[1]

HERMAN Y. EFRON

Veterans Administration Hospital, Lyons, New Jersey

Problem

This study attempts to determine whether for schizophrenic patients there would be any consistent changes in recognition threshold of tachistoscopically presented simple stimuli (capital letters) as a result of treatment with two phenothiazines (chlorpromazine and promazine) and phenobarbital. Primac, Mirsky and Rosvold [8] found that with normal subjects using the Continuous Performance Test there was a significant decrement following administration of chlorpromazine. Furthermore, using three dosage levels (100, 200, and 400 mg.), they found that higher dosages led to greater impairment than lower dosages. Similarly, using a wide variety of psychological tests and with normal subjects, Kornetsky, Humphries, and Evarts [2] found that secobarbital (100 and 200 mg.) and chlorpromazine (100, 200, and 400 mg.) significantly impaired performance. Again under higher dosage there was greater impairment than with lower dosages. When their subjects were required to discriminate between the size of two tachistoscopically presented circles the group receiving 200 mg. of secobarbital was more impaired than those given 200 mg. of chlorpromazine who in turn were more impaired than those given 100 mg. of secobarbital. On the other hand Lehmann and Hanrahan [4] using normal subjects found no decrement in reaction time, hand steadiness, digit span, and digit symbol under dosages of chlorpromazine "sufficiently large to produce somnolence." With 3 gr. secobarbital they found uniform impairment in all tested functions. Lehmann and Csank [3] using normal subjects found that chlorpromazine (100 - 150 mg.) produced a decrease in flicker fusion and tapping speed. All of the above experiments dealt with nonpsychiatric subjects and with performance after having been given a single or at most a few dosages of the drug under study.

Concerning the effects of chlorpromazine taken for a longer period of time Winter and Frederickson [12] found that it leads to "some disruption of coordination"

[1]This study was conducted as an individual hospital project of the Psychiatric Evaluation Project, Richard L. Jenkins, M. D., Director. It was done in conjunction with the Veterans Administration Cooperative Studies of Chemotherapy in Psychiatry Projects 1 and 2, which are described in detail elsewhere [10]. This paper was reviewed in the Veterans Administration and is published with the approval of the Chief Medical Director. The statements and conclusions of the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

particularly in the initial phases of treatment. They caution that ". . . evaluating performance on tests involving visual-motor coordination, *e.g.*, the Bender-Gestalt, account should be taken of concurrent drug therapy. However their study only covered a one week period and there was no gradual build-up to the maximum dosage of 50 mg. q.i.d. Using the maze test, Porteus[6] and Porteus and Barclay[7] found decrements in performance following treatment with chlorpromazine (300 mg. for six weeks to six months), but there have been at least two studies where these results were not duplicated[1, 5]. None of the previous studies included testing of recognition thresholds. The present study was designed to provide data on a basic psychological effect of prolonged treatment with the drugs utilized. Because of the empirical nature of the investigation the null hypothesis, that there are no differences between the effects of the various drugs, was formulated.

Method

Subjects. Forty nonlobotomized male patients, less than 50 years old and with a diagnosis of schizophrenia were the subjects in this study. The mean age was 39 ± 8 and the mean length of current hospitalization was 103 ± 104 months (Mdn = 109, Q_1 = 11, Q_3 = 144). No patient included was involved in individual or group psychotherapy, was receiving insulin or electroconvulsive treatment, or had any disease or illness which contraindicated the use of any of the drugs to be administered. Administrative decisions determined the number of patients from each ward to be included but after the quotas had been established patients were randomly selected from each of the wards.

Drugs. In order to vitiate the effects of any medication the patients were receiving at the time of selection, all tranquilizing medication was discontinued for at least one to two months prior to the initiation of the experimental medication. The drug schedule was: chlorpromazine 200 mg. b.i.d., promazine 200 mg. b.i.d., phenobarbital 1½ gr. b.i.d., or placebo b.i.d. To lessen the initial impact of the drugs there was a five day gradual build-up period before reaching maximum dosage. The drugs were supplied by the Central NP Research Unit of the Veterans Administration in identical capsules. No one at the hospital was aware of what drug each patient was receiving. After twelve weeks on the drugs there was a randomized cross-over design[2] in which two-thirds of the patients changed drugs and one-third were maintained on the same drug.

Apparatus. A constant illumination projection tachistoscope was used. Duration of exposure (10 milliseconds) and ambient illumination (70 footlamberts incident light at the screen and at the subject) were held constant and brightness was varied by a rheostat which supplied current to the lamp of the projector that contained the stimulus material. Initially the subject was presented with a trial series of six numbers in order to acclimate him to the task. The experimental series consisted of 10 capital letters projected onto a 21″ by 13″ beaded screen. Each number or letter was projected three times at the same brightness and if they were not correctly identified the rheostat was advanced two volts. Brightness level for the first presentation was the same for all subjects. When the subject identified a letter on two consecutive exposures the next letter was presented.

Testing Schedule. Recognition thresholds for the 10 stimulus letters were determined for each subject prior to initiation of the experimental medication, after 10-12 weeks on the first drug and for a third time 10-12 weeks after the randomized cross-over of medication.

Results and Discussion

Of the 40 patients, 27 were in sufficiently good contact and sufficiently cooperative to complete at least two of the three testing sessions. Comparisons were possible between a patient's performance after having been on the first drug, and after having been on the second drug with his baseline before receiving medication. If a patient received the same drug during both parts of the study only his performance at the end of the first 10-12 weeks was considered. Using this method of analysis, the number of comparisons between a subject's pre-drug performance and his performance after having been on one of the experimental drugs was as follows: chlorpromazine = 11, promazine = 11, phenobarbital = 9, placebo = 9.

The recognition threshold for each letter was taken as the mean of the rheostat settings (voltages) for the two consecutive correct identifications. For each testing session the median recognition threshold was then computed. Mean scores could not be obtained since there were a few instances where the subject recognized the letter correctly on the first two presentations and other instances where the threshold was

[2]For administrative reasons 8 of the 40 patients were discontinued after the first 12 week period.

so high its exact value was not determined. Using the difference between the median score of the pre-drug testing and the testing under the drug conditions the following comparisons were made: chlorpromazine *vs.* promazine, chlorpromazine *vs.* phenobarbital, chlorpromazine *vs.* placebo, promazine *vs.* phenobarbital, promazine *vs.* placebo, phenobarbital *vs.* placebo, chlorpromazine and promazine *vs.* phenobarbital and placebo, chlorpromazine and promazine *vs.* phenobarbital, and chlorpromazine and promazine *vs.* placebo. Statistical analysis by means of the Mann-Whitney U test[9, pp. 116-127] indicated that none of the differences were significant at the 5% level.

Our results indicate that for the drugs and dosages used there was no interference with the perceptual process tested, and suggesting that there is a difference between the initial effects of the drugs and their effects after prolonged usage. Our results are not consistent with the findings of Whitehead and Thune[11] who report greater cooperation in their chlorpromazine group. Of the four patients who cooperated for the initial testing and subsequently refused to cooperate, three were receiving chlorpromazine.

The fact that the phenobarbital group did not show any impairment is somewhat surprising in view of previous studies[2, 4]. Although they used a different barbiturate it would seem that 1½ gr. b.i.d. of phenobarbital would at least be equivalent to the dosage of Seconal (100 and 200 mg.) used by Kornetsky et al[2]. The most parsimonious explanation would seem to be that a tolerance is built up with prolonged usage of the drug.

Summary

Comparisons were made between the effects of chlorpromazine, promazine, and phenobarbital on schizophrenic patients' recognition thresholds of tachistoscopically presented simple stimuli (capital letters). Forty subjects were placed on one of the experimental drugs or a placebo and at the end of 12 weeks there was a randomized cross-over in which two-thirds of the patients changed medication. Results showed that there were no significant changes in recognition thresholds following treatment with the drugs. When viewed in conjunction with previous studies it suggests that impairment of functioning observed at the initiation of drug treatment may not persist when treatment is prolonged.

References

1. Daston, P. G. Stylus maze performance of chronic schizophrenics taking chlorpromazine. *J. consult. Psychol.*, 1958, 2, 384.
2. Kornetsky, C., Humphries, Ogretta and Evarts, E. V. Comparison of psychological effects of certain centrally acting drugs in man. *A. M. A. Arch. Neurol. Psychiat.*, 1957, 77, 318-324.
3. Lehmann, H. E. and Csank, J. Differential screening of phrenotropic agents in man: psychophysiologic test data. *J. clin. exp. Psychopathol.*, 1957, *18*, 222-235.
4. Lehmann, H. E. and Hanrahan, G. E. Chlorpromazine: new inhibiting agent for psychomotor excitement and manic states. *A. M. A. Arch. Neurol. Psychiat.*, 1954, 71, 227-237.
5. Mason-Browne, N. L. and Borthwick, J. W. Effect of perphenazine (Trilafon) on modification of crude consciousness. *Dis. Nerv. System*, 1957, 18, 300-308.
6. Porteus, S. D. Maze test reactions after chlorpromazine. *J. consult. Psychol.*, 1957, *21*, 15-21.
7. Porteus, S. D. and Barclay, J. E. A further note on chlorpromazine. *J. consult. Psychol.*, 1957, *21*, 297-299.
8. Primac, D. W., Mirsky, A. F. and Rosvold, H. E. Effects of centrally acting drugs on two tests of brain damage. *A. M. A. Arch. Neurol. Psychiat.*, 1957, 77, 328-332.
9. Siegel, S. *Non-parametric statistics.* New York: McGraw-Hill, 1956.
10. *Transactions of the second research conference on chemotherapy in psychiatry.* Vol. 2. Washington: Veterans Administration, 1958.
11. Whitehead, W. A. and Thune, L. E. The effects of chlorpromazine on learning in chronic psychotics. *J. consult. Psychol.*, 1958, *22*, 379-383.
12. Winter, W. D. and Frederickson, W. K. The short-term effects of chlorpromazine on psychiatric patients. *J. consult. Psychol.*, 1956, *20*, 431-434.

EFFECTS OF MORPHINE AND PENTOBARBITAL ON DIFFERENTIAL MMPI PROFILES

CHARLES A. HAERTZEN AND HARRIS E. HILL

National Institute of Mental Health, Addiction Research Center, PHS, Lexington, Kentucky

PROBLEM

Evidence that altering certain incentives modifies the behavioral effects of morphine and pentobarbital was presented in a previous report.[6] The present study was similar in formulation but was specifically concerned with possible correlations between drug effects and personality characteristics as measured by the MMPI. That this inventory may be useful in assessing drug effects has been demonstrated by Belleville in testing LSD-25,[2] and by Isbell, *et al*,[8] and Abramson[1] in using alcohol. No mention was made in these reports of a possible interaction between personality and drug effect. Some support for this contention, however, is supplied by the studies of Lasagna and Beecher[9] and by Felsinger, *et al*,[3] who reported that the 7 of their group of 20 normal subjects who "responded unequivocally with euphoria to heroin or morphine, or both", were the most immature of the group as well as being self-centered, impatient, and above the median in "anxiety, hostility content".

These studies demonstrated that individuals within a "normal" population may respond differently to the same drug. Furthermore, clinical observation suggests that even within a relatively homogeneous group of drug adicts differential effects of addicting drugs may be expected. The purpose of the present study, then, was to investigate systematically the effects of acute administration of morphine and pentobarbital on responses to the MMPI, and, within the limits of this instrument to assess the degree to which drug effects are dependent upon personality characteristics.

METHOD

Subjects. Fifty-five former addicts serving sentences for violation of the Harrison Narcotic Act volunteered for the study. Approximately equal numbers of white and negro subjects were used. The group form of the MMPI was given in balanced order under control, morphine, and pentobarbital conditions. Only one test per week was given to avoid residual drug effects. Two intramuscular injections of morphine (30 mg. each) or of pentobarbital (100 mg. each) were given (9:30 a. m. and 1:30 p. m.). The MMPI was administered at 2:30 p. m. the same day. Standard instructions were given for completing the inventory, except that the subjects were told to respond according to the way they felt on the day of the test. An informal atmosphere was encouraged and the men were allowed to smoke and make comments to each other.

Present personality groupings were based upon a "high point" classification used for a previous investigation in which 94 per cent of the population of addicts presented abnormal profiles under drug-free conditions.[7] To distinguish subsamples in this prior study, all abnormal profiles were classified according to the two most elevated scales as falling into one of the following subgroups: (a) *Psychopathic group:* Profiles having Pd and Ma as the highest elevations; (b) *Neurotic group:* Profiles having two highest points among the Hs, D, Hy, and Pt scales; (c) *Schizoid group:* Profiles having two highest points among the Pa, Sc, Ma, and Pt scales; (d) *unclassifiable group:* Profiles that could not be classified by such high-point criteria [The Pd scale was disregarded except for group (a)]. The generic term "Conduct Disorder", indicating variation in the direction of psychopathy, was applied to these four groups because the Pd scale was elevated approximately two standard deviations in all these subsamples. For present purposes in classifying drug addicts, however, the terms "Conduct Disorder, Drug Addict," with the various subgroupings, have been abbreviated to: Primary psychopath, Neurotic psychopath, Schizoid psychopath, and Unclassifiable group.

Procedure. In the present study the eight clinical scales were ranked for each profile, using Hathaway's[4] coding system. Rank order correlations were then computed between each subject's control MMPI and the mean rank orders of the three classes of addict profiles obtained previously.[7] Percentile ranks of correlation coefficients for all subjects within each classification of addicts were then established. In this way all individuals were ranked for similarity of profile in each of the three groups, and each subject was ordered to the group for which his percentile rank was the highest. In case of a tie on percentile rank, the decision for placement was based on the size of the correlation coefficient. If *rho* was less than .45, the profile was placed in an unclassifiable group; as in the previous study, most of the profiles in this group were abnormally elevated

but could not be classified according to present criteria. Taylor's[11] Manifest Anxiety Scale (A), Welsh's[12] Anxiety Index (A1), and Internalization Ratio (IR) were also used. The Masculinity-femininity (Mf) scale was not employed.

A mixed analysis of variance design was used to test the effects of personality groupings, scales, and drugs, and independent analyses of the same type were used to test changes in A, AI, and IR, as well as in F + Pd, a combination score designed to measure interpersonal behavior which is generally recognized as being undesirable. A treatments × treatments × subjects design was utilized to test drug and scale effects within each group.[10]

Results

The means for all MMPI scales and indices for each of the subgroups are presented in Table 1. Application of analysis of variance revealed significant effect due

Table 1. Summary of Means for All Drug and Nondrug Conditions

	L	F	K	Hs	D	Hy	Pd	Pa	Pt	Sc	Ma	AI	IR
Primary Psychopath													
Control	45	57	52	45	48	50	69	45	44	52	71	45	.72
Morphine	45	63	52	52*	54	53	73	51*	55**	60*	70	58*	.82**
Pentobarbital	46	61	53	50	54	53	75	48	49†	56	71	52	.77†
Neurotic Psychopath													
Control	49	61	53	66	79	63	73	54	66	60	56	86	1.10
Morphine	48	56	54	63	70*	60	70	56	62	60	62	73*	1.00**
Pentobarbital	48	62	51	63	72	60	70	59	64	67	64	77	1.03*
Schizoid Psychopath													
Control	46	68	48	50	56	49	63	60	60	72	70	68	.92
Morphine	45	71	48	57	59	52	70	61	60	75	78	66	.88
Pentobarbital	46	66	48	52	56	51	69	59	59	69	75	65	.85*

* Significantly different from control at <.05 level
** Significantly different from control at <.01 level
† Significantly different from morphine at <.05 level

to groups, scales, drugs × groups, as well as drugs × scales × groups interactions.[1] However, it is noteworthy that no significant general effect of drugs was found.

The drugs, scales, and drugs × scales variances were significant for the Primary group. The significant drug effect for this group accounts for the drugs × groups interaction found in the analysis of the total sample, since drug effects alone were not significant for the Neurotic or the Schizoid groups. The drugs × scales effect in the Primary group was further traced to significant increases in the HS, Pa, Pt, and Sc scales from control to morphine conditions. The drugs × scales effect in the Neurotic subgroup of psychopaths was limited to a significant decrease in the Depression scale under morphine. No significant drug or drugs × scale effects were found in the Schizoid or Unclassified groups.

The Anxiety Index and the Internalization Ratio were analyzed separately. As shown in Table 1, both were significantly decreased by morphine in the Neurotic group and both were significantly increased by this drug in the Primary psychopath group. Under pentobarbital, both the Neurotic and Schizoid groups showed significant decreases on IR. The Primary psychopath group was significantly higher on IR under morphine than under pentobarbital. Analysis of the F + Pd scores is not presented in tabular form. Increases in this measure were significant beyond the .01 per cent level from control to both morphine and pentobarbital in the Primary psychopath, but the obtained decreases in the Neurotic group were nonsignificant.

Judging from the subjects' behavior, the dose of morphine was higher than that required to produce a measurable effect on the MMPI. In spite of the large dose all completed the test, and although some vomited, this effect was described in the addicts' jargon as a "good sick". On the other hand, the dose of pentobarbital was somewhat lower than could have been used for present purposes.

[1]The table of analysis of variance can be obtained from the authors on request.

Discussion

All drug effects obtained were dependent upon personality classification; no significant changes on the MMPI scales were produced by the drugs before the subjects were classified into personality groupings. It would thus seem that combining test results from even moderately heterogenous individuals may obscure some psychological drug effects.

Significant decreases produced by morphine on the Anxiety Index and Internalization Ratio in the Neurotic psychopath group is evidence, complementary to clinical observations and experimental findings, that opiates reduce fear and anxiety. The decrease in elevation of the Depression scale is similar evidence, and adds considerable credence from an operationally defined source, that certain subsamples of psychopaths and addicts have a strong depressive trend that opiates are effective in ameliorating.

The only significant change found in the Schizoid psychopath group was a reduction in the Internalization Ratio produced by pentobarbital. This is in accord with clinical observation and general belief that barbiturates relieve some forms of anxiety and enhance "externalization" of conflicts. However, the lack of significant change on other indices and scales may have been due to the amount of drug given, the size of the sample, degree of sensitivity of the MMPI, or it may have been due to heterogeneity of error variance within the group.

The increased elevation induced by morphine on the Hs, Pa, Pt, and Sc scales in the Primary psychopath group must be interpreted with caution. Since all scales except Pd and Ma were within the normal range under control conditions, as was found in a previous study,[7] this group would not be considered either anxious or depressed. It would not be expected, then, that drugs would produce a decrease in the neurotic triad or indices of anxiety in the Primary group as was found for the Neurotic psychopath. Two alternative explanations, however, are suggested for the significant increases that were found. Morphine appears either to increase symptoms in the Primary psychopath which are represented by the altered scales and indices, or to change the individual so that he is more willing to admit such symptoms and behavior.

Until direct tests of these hypotheses are made, the latter possibility seems the more reasonable in view of three lines of suggestive evidence: (a) Astin,[2] applying factor analysis to the Pd scale obtained on a group of addicts who were maintained in a drug-free environment, reported that the Primary psychopath "responds in over-perfect ways to the MMPI items and denies any weakness in interpersonal relations;" (b) clinical experience indicates that morphine reduces concern in interpersonal relations; (c) the combination F + Pd, utilized in the present study as a measure of social behavior generally recognized as being undesirable, was found to be significantly increased in the Primary but not in the Neurotic psychopath. The evidence is thus suggestive that the above-mentioned tendency toward "denial" in the Primary group is reduced by opiates and that this alteration in motivation is reflected in the endorsement of MMPI items.

Summary

The MMPI was given in balanced order to 55 postaddict subjects under control, morphine (60 mg.), and pentobarbital (200 mg.). Drug effects were assessed for the total group as well as for four subgroups: Primary psychopath, (N = 12); Neurotic psychopath, (N = 12); Schizoid psychopath, (N = 12); and Unclassifiable, (N = 19). The drugs did not alter scale elevations for the combined group; significant drug effects were dependent upon differential personality characteristics as measured by the MMPI. The only significant change under pentobarbital was a decrease in the Internalization Ratio in the Schizoid and Neurotic groups. Morphine produced unpredicted increases in Hs, Pa, Pt, Sc, and in the Internalization Ratio in

[2]Astin, A. A factor study of the MMPI Psychopathic deviate scale. *J. consult. Psychol.*, in press.

the Primary psychopath group, and produced predicted decreases in D, the Anxiety Index, and Internalization Ratio in the Neurotic psychopath. The latter changes were anticipated since it had been hypothesized that one important aspect in the development of addiction to opiates is initial reduction of discomfort, anxiety, and depression in those addicts who show neurotic tendencies. Although the results suggest that opiates reduce sensitivity to interpersonal relations in the Primary psychopath, the implications of the findings for addiction liability of the opiates in this type of individual are not as yet clearly defined.

References

1. Abramson, H. A. The effect of alcohol on the Minnesota Multiphasic Personality Inventory. *Ann. N. Y. Acad. Sci.*, 1946, *46*, 535-557.
2. Belleville, R. E. MMPI score changes induced by Lysergic Acid Diethylamide. *J. clin. Psychol.*, 1956, *12*, 279-282.
3. Felsinger, J. M. von, Lasagna, L., and Beecher, H. K. Drug induced mood changes in man. 2. Personality and reactions to drugs. *J. Amer. Med. Assoc.*, 1955, *157*, 1113-1119.
4. Hathaway, S. R. A coding system for MMPI profiles. *J. consult. Psychol.*, 1947, *11*, 334-337.
5. Hathaway, S. R., and McKinley, J. C. *Minnesota Multiphasic Personality Inventory Manual* (Revised ed). New York: The Psychological Corp., 1951.
6. Hill, H. E., Belleville, R. E., and Wikler, A. Motivational determinants in modification of behavior by morphine and pentobarbital. *A. M. A. Arch. Neurol. Psychiat.*, 1957, *77*, 28-35.
7. Hill, H. E., Haertzen, C. A. and Glaser, R. Personality characteristics of narcotic addicts as indicated by the MMPI. *J. gen. Psychol.*, in press.
8. Isbell, H., Fraser, H. F., Wikler, A., Belleville, R. E. and Eisenman, Anna J. An experimental study of the etiology of "Rum Fits" and delirium tremens. *Quart. J. Stud. Alcohol*, 1955, *16*, 1-33.
9. Lasagna, L., Felsinger, J. M. von, and Beecher, H. K. Drug induced mood changes in man. 1. Observations on healthy subjects, chronically ill patients, and "Postaddicts." *J. Amer. Med. Assoc.*, 1955, *157*, 1006-1020.
10. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton-Mifflin, 1953.
11. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, *48*, 285-290.
12. Welsh, G. S. An anxiety index and internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, *16*, 65-72.

INTELLECTUAL CHANGES FOLLOWING FRONTAL LOBE PROCAINIZATION[1]

ALBERT F. PAOLINO AND IRA FRIEDMAN

Cleveland Psychiatric Institute and Hospital

Problem

Post-operative intellectual changes following psychosurgery have been reported by various investigators[1, 2, 3, 4]. The present study is addressed to the problem of intellectual changes following the psychosurgical (pharmacosurgical) technique of injecting procaine into the frontal lobes[5].

Procedure

The sample consisted of 26 patients hospitalized at the Summit County Receiving Hospital who received an injection of 20 to 30 cc. of 0.5 per cent procaine solution into each frontal lobe in the plane of the coronal suture. A control group of 26 patients hospitalized at the Cleveland Psychiatric Institute and Hospital was matched for age, sex, initial IQ level, and diagnosis. Most of the patients were diagnosed as schizophrenic, 17 in the experimental and 18 in the control group. The

[1]Modification of a paper presented at the meeting of the Ohio Psychological Association, Cleveland, Ohio, November 1, 1958.

mean age for operatees was 37.73 years, S.D. of 10.55 and the mean age for the control group was 37.61, S.D. of 9.81. The mean initial IQ level was 99.92, S.D. of 14.19, for the operatees and 99.73, S.D. of 14.96, for the control group. The control group received no psychiatric treatment other than routine intake, admission, and diagnostic interviews.[2]

The Wechsler Bellevue Intelligence Scale was administered twice to each subject. In the experimental group, approximately 10 days elapsed between the first and second testing, and the second testing was completed within 48 hours after procaine injection. The control group received its second test administration from one to three weeks following the first. Three sequences of intelligence test administration were employed with the two groups; Form I of the Wechsler Bellevue Intelligence Scale for both initial and final testing with 11 patients in each group, Form I then Form II with 8 patients in each group, and Form II then Form I with 7 patients in each group.

The verbal, performance, and full scale IQs of the two groups were analyzed by means of the Mann-Whitney U-test, a non-parametric technique. The *t*-test was also used but it is of doubtful validity due to the skewed nature of the data or its lack of homogeneity of variance.

Results and Discussion

Table 1 presents the mean IQ changes of the experimental and control groups along with the significance of these changes within each group from first to second testing, as well as the significance of the changes between the experimental and control groups.[3] The results indicate that both groups reveal highly significant increases in full scale IQ scores. These increases appear to be more related to increases in performance or non-verbal functioning than to increases in the verbal area. In addition, the differences tend to be greater and more striking in the experimental than in the control group. However, when the changes in the two groups are compared, the differences are not significant with the exception of two borderline .05

Table 1. Significance of the IQ Changes Within Experimental and Control Groups and Between Experimental and Control Groups

Test Order	Experimental Mean Change	*t*	*p**	*p***	Control Mean Change	*t*	*p**	*p***	Between Groups *t*	*p**	*p***
Full Scale IQ											
Total	+4.88	5.22	.001	.001	+3.12	2.80	.01	.01	1.32	NS	NS
I - I	+5.82	4.23	.01	.01	+1.64	1.15	NS	NS	2.11	.05	NS
II - I	+5.63	2.88	.05	.05	+5.88	2.62	.05	.05	−.11	NS	NS
I - II	+2.57	1.62	NS	NS	+2.29	1.02	NS	NS	.10	NS	NS
Verbal Scale IQ											
Total	2.23	1.89	NS	.05	1.46	.90	NS	NS	.36	NS	NS
I - I	2.55	2.22	NS	NS	−.45	.42	NS	NS	1.28	NS	NS
II - I	8.13	2.82	.05	.01	4.50	1.03	NS	NS	.77	NS	NS
I - II	−5.00	−4.82	−.01	−.01	1.00	.46	NS	NS	−2.08	NS	NS
Performance Scale IQ											
Total	7.62	6.20	.001	.001	3.77	2.86	.01	.01	2.14	.05	NS
I - I	8.45	6.39	.001	.01	4.09	1.81	NS	NS	1.66	NS	NS
II - I	5.00	1.99	NS	.05	4.88	2.01	NS	NS	.04	NS	NS
I - II	9.29	3.19	.05	.05	2.00	.92	NS	NS	2.00	NS	NS

*McNemar's Method of Differences
**Non-parametric analysis, Wilcoxon Matched Pairs Signed Rank Test

[2]Five patients in the control group received a placebo taken orally in connection with another research project.

[3]The analysis of individual subtests is not presented. There were no significant differences in any of the subtest changes between experimental and control groups. This finding is in essential agreement with those of King[3] although his controls showed significant improvement in Object Assembly at the second post-testing period.

p-values suggesting greater increases in the experimental group in performance IQ and greater increases in full scale IQ when identical tests are employed.

These results indicate clearly that intellectual functioning immediately following frontal lobe procainization reveals no evidence of deterioration. The only suggestion of intellectual impairment observed was in the 7 experimental subjects who displayed a mean verbal IQ loss of 5 points.[4] These same subjects, however, had a performance IQ gain of 9.29 points and a full scale IQ increase of 2.57 points.

Summary and Conclusions

This study investigates intellectual changes following the psychosurgical (pharmacosurgical) technique of injection of procaine into the frontal lobes. A group of 26 operatees was tested with the Wechsler Bellevue Intelligence Scale before and after the operation and was compared with a control group matched as to age, sex, initial IQ and diagnosis. The results indicate that no intellectual impairment occurs immediately following this operation, that significant practice effects are in evidence similar to those revealed by a matched control group, and that there is some suggestion that the operative procedure may result in greater increases in performance IQ than those attributable to practice effects.

References

1. Dubin, S. S. and Thaler, M. The use of psychological tests on schizophrenic patients before and after shock treatment. *Amer. Psychologist*, 1947, *2*, 283. Abstract.
2. Kessler, Lucille. Intellectual changes in schizophrenic patients following electric shock therapy. *Amer. Psychologist*, 1948, *3*, 277-278.
3. King, H. Intellectual function. In F. A. Mettler, (Ed.) *Selective Partial Ablation of the Frontal Cortex*. New York: Paul B. Hoeber, 1949.
4. Porteus, S. D. and Peters, H. N. Maze test validation and psychosurgery. *Genet. Psychol. Monogr.*, 1947, *36*.
5. Whitworth, J. M., Weinberger, L. M., Zimbler, M. B., and Paolino, A. F. The use of frontal lobe procaine injections in the treatment of mental disorder. *Ohio St. med. J.*, 1956, *52*, 727-728.

[4]This group of subjects was younger and less bright than the other experimental subjects. Further comparisons, not presented here, indicated trends which suggest that the greatest improvement in intellectual functioning concomitant with this operative procedure may occur in brighter and older patients who, in addition to performance increases, show verbal increases as well.

ANXIETY AND SOCIO-ECONOMIC STRATIFICATION

ERWIN J. LOTSOF AND RICHARD CENTERS

University of California, Los Angeles

Introduction

Our present-day American culture has been termed by some insightful observer "the age of anxiety" and, indeed, the study of and concern with anxiety as a psychological variable is no longer confined to the clinician or psychopathologist, but enters increasingly into the sphere of interest of the social psychologist. Probably contributing to this expanding concern is the relatively great amount of attention given in the various mass media of communication to tranquilizing drugs purported to relieve anxiety and tension. Several months ago, in the Los Angeles area, it was even rather common to see large signs on drug store windows indicating that these drugs were in stock, and numerous jokes and quips on radio, television and nightclub shows further suggested the widespread interest and use of these anxiety relievers. Such a public climate of interest and concern with anxiety suggested the feasibility of testing certain hypotheses relating the anxiety factors to other variables by means of a

survey approach. Several previous investigations[2, 4] have suggested that anxiety and somatization reactions are related to socio-economic stratification, but in all cases these studies have been based upon highly selected samples, *i.e.*, clinical patients, and it was desired to test this relationship with a more adequate sample.

On the basis of observations and reports in the literature, it was hypothesized that individuals in the lower socio-economic strata should show more concern over somatic factors than individuals in the higher socio-economic groups. Further, it was felt that the higher the educational achievement the fewer should be the somatic complaints. Also, it was hypothesized that a sense of personal competence should be inversely related to concern with somatic factors.

Method

A questionnaire was constructed to assess the degrees of anxiety, tension, concern with soma, and personal competence. Some of our anxiety items were derived from the Minnesota Multiphasic Personality Inventory, and the measure of feelings of personal competence was adapted from a study by Douvan and Walker[1]. The interview schedule also contained questions concerned with socio-economic status, educational and other background factors.

The data for this research are the results of person-person interviews with a sample of 1,250 persons, constituting a cross section of the adult population of the Los Angeles Metropolitan area. Each person was interviewed alone at his home. Both sexes, above 18 years of age, of various racial and ethnic derivation were included.

The replies to the interview were coded, employing a dichotomizing procedure, and *phi* correlations were obtained for each of the 44 variables encompassed. A factor analysis using Kaiser's Varimax method[3] was performed on the data. This method is an orthogonal iterative procedure which tends to maximize the variance of the squared extended vector projections, giving a solution similar to simple structure.[1] In what follows only variables showing loadings of approximately .30 and above will be given importance. The variable of .30 is arbitrary, since the standard error is not known.

Results and Discussion

A total of fourteen factors were identified, but only seven of these seem sufficiently important to warrant description and comment, the other seven tending to have fewer loadings on them with relatively lower intercorrelations of the items encompassed.

Factor 1 "Anxiety" loads mainly on questions pertaining to anxiety, tension and concern with physical well-being or bodily health. For example, some of the items involved and their loadings are as follows:

> "During the past year, how often have you experienced periods of a rapid heart or pulse; very often, fairly often, or hardly at all? (very often and fairly often were scored plus, hardly at all was scored minus)." Loading .48.
>
> "During the past year, how often have you had times when you felt all worn out or 'dead tired'; very often, fairly often, or hardly at all?" (Very often and fairly often were scored plus, hardly at all was scored minus) Loading .46.
>
> "During the past year, how often have you been worried about your health; very often, fairly often, or hardly at all?" (Very often and fairly often were scored plus, hardly at all was scored minus) Loading .58.

Our prediction that the anxiety factor should also load on items concerned with socio-economic variables was not borne out in our study.

The second factor, which seems to be mainly socio-economic in nature, loads on rated socio-economic status (average and above were scored plus, below average

[1]The complete interview schedule, correlation matrix and rotated factor matrix are on file with the American Documentation Institute, Washington, D. C. Order Document 5866, remitting $1.25 for photoprints, or 35 mm. microfilm. Make checks payable to Chief, Photo-duplication Service, Library of Congress, Washington 25, D. C.

scored minus); house type rating (excellent, very good and good scored plus; average, fair, poor, and very poor scored minus). Dwelling area rating (scored like the preceding item). Occupation (white collar occupation were scored plus, manual work occupations were scored minus). Class identification (upper and middle class identifications were scored plus, working and lower class identifications were scored minus). There was a moderately high loading of this factor on the item: "Have you heard or read about certain drugs or medicines called tranquilizers?" (A "yes" answer was scored plus) Possibly it is due simply to the fact that higher socio-economic status persons have more commonly become informed of ataractic drugs than those of lower status.

Factor 3, "religiosity", loads mainly on questions asking whether the respondent was a member of a church ("yes" was scored plus); which denomination (Christian scored plus, Jewish scored minus), and whether he considered himself devout (scored plus) or not devout (scored minus). It had been supposed that religiosity would be positively related to anxiety. However, it appears not to be so to any appreciable degree, nor is it appreciably related to any of the other variables encompassed in our study.

Another factor, number four "feelings of inadequacy" is similar to the proceding one in its isolation from other variables. It loads mainly on the following three items.

> "During the past year, how often have you felt very blue or depressed; very often, fairly often, or hardly at all?" (The "often" responses were scored plus)
>
> "Problems and difficulties are always coming up in peoples' lives. To what extent have you felt in this past year that problems and difficulties were so great that you might not overcome them; very often, fairly often, or hardly ever?" (The "often" responses were scored plus)
>
> "People often have feelings that they are not as good as or are inferior to other people; to what extent in this past year have you felt this way; very often, fairly often, or hardly at all" (The "often" responses were scored plus)

The fifth factor "Information and Experience" appears to load on having heard or read about medicines called tranquilizers, educational level achieved and the age of the individual. The factor does not appear to be related to anxiety or other variables to an appreciable degree.

Factor 6, "Self-confidence" loads on the following three questions: "Once you decide to do something, how sure are you of the outcome?" ("Very sure" and "fairly sure" were scored plus) "How sure are you usually that your opinions and beliefs are right? ("Very sure" and "fairly sure" were scored plus) and, "to what extent do you feel that you will be able to get the things you want out of life—very confident, fairly confident, or not at all confident." ("Very confident" and "fairly confident" were scored plus) It had been our expectation that there would be an inverse relationship between feelings of self-confidence and indicators of tension and anxiety, but such did not turn out to be the case.

The seventh factor loads on items relating to feelings of self-competence or optimism: "Some people feel that their lives are worked out just the way they wanted; others feel that they have really had bad breaks; how do you feel about the way your life is turning out—satisfied or dissatisfied?" ("Satisfied" was scored plus) "Some people feel that they can make pretty definite plans for their lives for the next few years; others feel they aren't in a position to plan ahead. How about you—do you feel able to plan ahead or not?" ("Able to plan ahead" was scored plus) "What do you think your chances are of living the kind of life you'd like to have? Do you think they are pretty good, or not so good?" ("Pretty good" was scored plus). We hypothesized initially that these questions which comprise the index of personal competence feelings and which concern optimism would load in an opposite direction to the anxiety items and in the same direction as self-confidence. Our hypothesis again was not borne out by the data. This finding may cast some doubt that feelings of self-confidence and optimism are positively correlated to each other and negatively correlated with tension and anxiety.

SUMMARY

This study concerned itself with anxiety and socio-economic stratification, both assessed by a questionnaire administered in person-to-person interviews conducted with a cross section of the adult population of the Los Angeles metropolitan area. Correlations were obtained for the variables and then factor analysed, using Kaiser's Varimax method. Although the anxiety items tend to load on the same factor, Factor 1, they do not appear to be related to socio-economic status. A rather interesting finding was that optimism and self-competence were not inversely related to anxiety. It would appear that, on the basis of the present study, anxiety and somatization reactions are unrelated to socio-economic status in the general population.

REFERENCES

1. DOUVAN, ELIZABETH, and WALKER, ALAN M. "The Sense of Effectiveness in Public Affairs." *Psychological Monographs*, 1956, *70*, No. 22.
2. HOLLINGSHEAD, A. B., and REDLICH, F. C. *Social Class and Mental Illness.* New York: John Wiley and Company, 1958.
3. KAISER, HENRY F. "An Analytic Rotational Criterion for Factor Analysis." A paper presented at the Sept. 1955 meeting of the American Psychological Association.
4. RUESCH, J., "Social Technique, Social Status, and Social Change in Illness" in Kluckhohn, C., Murray, H. A., and Schneider, D. N. *Personality in Nature Society and Culture.* New York: Knopf, 1953.

A STUDY OF FACTORS RELATED TO LENGTH OF STAY IN PSYCHOTHERAPY*

M. A. BAILEY, L. WARSHAW AND R. M. EICHLER

Veterans Administration, New York Regional Office

PROBLEM

A previous investigation by the authors[1] revealed that the patients applying for treatment in the Mental Hygiene Clinic of the New York Regional Office are representative of the veteran population in the greater New York area for most demographic characteristics. However, assignment to psychotherapy was found to be related to normative data indicative of high socio-economic status, intrapsychic complaints, youth, expressed desire for psychotherapy, recommendations from psychological test evaluation, and previous treatment. Conversely, assignment to the psychosomatic clinic was found to be related to low socio-economic status, somatic complaints, requests for medication, lack of recommendation from psychological test evaluation, and no previous psychotherapy. The present investigation is designed to determine whether the above and other demographic criteria for assignment are related to length of stay in psychotherapy or in psychosomatic treatment.

METHOD

Items of information were recorded on a personal data card in the process of screening patients to provide data relevant to: age, number of years of schooling, occupation, religious background, race, diagnosis by psychologists and therapists, nature of complaint, treatment requested by patient, treatment recommended by the psychologist following psycho-diagnostic testing, and previous treatment experience. Approximately one year after the date of testing, the treatment folders of these patients were reviewed to determine (a) length of stay in treatment, and (b)

*The authors wish to express their appreciation to Mrs. Louisa Smith for assistance in collecting the data and to Drs. R. H. Paynter and Leon Rostker for many helpful suggestions.

amount of improvement. Amount of improvement was determined from the therapists' evaluations in their progress and summary notes. It was possible to utilize length of stay in treatment as a criterion of improvement because of the very significant relationship obtained between the two, considerably beyond .01 (chi square 61.23). Length of stay was divided into the following categories: two months or less, three to six months, seven to eleven months, and twelve or more months. This afforded the investigators a quantitative means of evaluating the effectiveness of treatment. Patients were divided into two treatment groups: one group consisted of 247 psychosomatic patients, and a second group consisted of 211 psychotherapy patients. Normative data were then studied in relation to length of stay in treatment for each group. Chi squares and Pearson Product Moment correlations were computed in order to study the above relationships.

Results

Psychosomatic Patients. Chi square tests did not reveal any significant relationships between length of stay in treatment by medication and any of the demographic or clinical data studied. However, trends were noted in the relationship between length of stay in psychosomatic therapy and level of occupation ($p = 19$), diagnosis by psychologists ($p = .10$), diagnosis by the therapists ($p = 15$), treatment requested by the patient ($p = .08$), treatment recommended by the psychologist ($p = .20$). These relationships indicate that patients in lower level occupations, patients with psychotic diagnoses, patients requesting medication, and patients recommended for medication subsequent to psychological testing tend to continue treatment by medication for longer periods of time than those not presenting the above characteristics.

Psychotherapy Patients. Chi square tests did not reveal any significant relationships between length of stay in psychotherapy and occupational level, religious background, diagnoses by psychologists and therapists, nature of complaint, treatment requested or psychologists' recommendations.

Significant positive relationships were found between length of stay in psychotherapy and number of years of schooling (chi square 32.19, $p < .01$) and having had previous experience in psychotherapy (chi square 11.95, $p < .01$). There was also a small but significant correlation between the length of stay in previous psychotherapy and the length of stay in subsequent psychotherapy (N 118, r .25, $p = .01$).

Discussion

In a previous study by the investigators[2] it was found that criteria such as relative youth, high educational and occupational level, intrapsychic complaints, expressed interest in a particular form of treatment, psychologists' test recommendations, and previous treatment experience played a significant role in determining the type of treatment to which patients were assigned. The present findings strongly suggest that most of the criteria utilized for the selection of psychotherapy patients are not valid in predicting their length of stay in psychotherapy with the exception of number of years of schooling, and previous treatment experience. Similarly, the criteria utilized in selecting patients for treatment in the Psychosomatic Clinic are not valid in determining their length of stay in treatment. However, some trends are noted in predicting length of stay of psychosomatic patients in treatment which suggest that theoretical assumptions such as the use of medication for psychotic patients, patients with fewer years of schooling, and patients who are somatically oriented have some pragmatic value.

Summary

Personal data cards and treatment folders of 247 psychosomatic patients and 211 psychotherapy patients were reviewed to study the relationship between certain normative data and length of stay in treatment. A chi square test revealed a highly significant relationship between improvement in psychotherapy and length of stay

in treatment. Although no statistically significant relationships were obtained between normative data and length of stay in treatment in the psychosomatic clinic, several trends were found which established some relationship between length of stay in treatment and low socio-economic status, diagnosis of psychosis, and somatic orientation. Length of stay in psychotherapy was found to be significantly related to number of years of schooling, and amount of previous experience in psychotherapy. On the basis of these findings it appears that most criteria for assignment used in this clinic are not valid in predicting the length of stay in treatment.

References

1. Bailey, M. A., Warshaw, L. and Eichler, R. M. A survey of patients screened in a mental hygiene clinic. Unpublished Manuscript. Medical Library, Veterans Administration, New York Regional Office, N. Y. 1958.
2. Bailey, M. A., Warshaw, L. and Eichler, R. M. A study of criteria for selecting psychotherapy cases in a mental hygiene clinic. Unpublished Manuscript. Medical Library, Veterans Administration, New York Regional Office, N. Y. 1958.

MULTIDISCIPLINARY VIEWS ON THE PREPARATION OF WRITTEN CLINICAL PSYCHOLOGICAL REPORTS: III. THE TROUBLE WITH PSYCHOLOGICAL REPORTS

NORMAN TALLENT
Veterans Administration Center
The George Washington University

WILLIAM J. REISS
Veterans Administration Center

Introduction

Many clinical psychologists and their colleagues on the psychiatric team appear to share a feeling that the psychologist's written clinical evaluation is often not as effective a contribution to patient care as it might be. To the end of determining what improvements should perhaps be made, an extensive survey of opinion about the negative features of current psychological reports was conducted.

Procedure

Source of Data. An incomplete sentence (or paragraph) root reading "The trouble with psychological reports is . . .," preceded by the instruction "Please complete in as few or as many words as you see fit:" was presented as an "optional" exercise on a research form used for another study in the area of clinical psychological reporting. This form was mailed to 741 psychiatrists, 433 psychiatric social workers, and 393 psychologists, all Veterans Administration staff persons. Completed forms were received from 606 psychiatrists, 384 psychologists, and 421 social workers, but of these only 276 psychiatrists, 233 psychologists, and 203 social workers completed the sentence (or more usually, the paragraph) root.

Categorization of Data. The ordering of the raw data into some sort of categorized presentation (see Table 1) was a prodigious task, the transcribed responses numbering 118 pages, single spaced. Differences in manner of expression from respondent to respondent at times made it difficult to know where several responses represented essentially the same idea, *i.e.*, properly belonged in the same category, or contained some important variations. Nevertheless, it was possible for the investigators jointly to consider one response at a time and to translate it into some meaningful expression. For example, the statement "The trouble with psychological reports is . . . they are attempts on the part of the examiner to show what *he can do* rather than what the

TABLE 1. CRITICISMS OF PSYCHOLOGICAL REPORTS*

		Psychiatrists	Psychologists	Social Workers
I.	PROBLEMS OF CONTENT			
	A. Unnecessary duplication	< 1	< 1	< 1
	B. Minor relevancy	5	5	2
	C. Offer diagnosis	8	< 1	< 1
	D. Offer prognosis	4	0	0
	E. Offer recommendations	8	< 1	< 1
	F. Improper emphasis	10	13	10
	G. Too much raw data	5	12	5
	H. Too little raw data	20	4	8
	I. Delete essential material	5	13	7
II.	PROBLEMS OF INTERPRETATION			
	A. Too general (reports do not differentiate among patients)	9	30	9
	B. Too speculative	13	6	3
	C. Speculation not labeled as such	3	7	1
	D. Assorted irresponsible interpretations	41	27	22
III.	PROBLEMS OF PSYCHOLOGIST'S ATTITUDE OR ORIENTATION			
	A. Overabstract	3	2	0
	B. Too authoritative	4	2	2
	C. Exhibitionistic	4	6	2
	D. Not patient oriented	0	3	< 1
	E. Test oriented	0	5	< 1
	F. Too theoretical	2	3	2
	G. Not written with practical or useful purpose in mind	8	41	16
	H. Miscellaneous (Especially: too intellectual, lack of interest in communication)	1	5	2
IV.	PROBLEMS OF COMMUNICATION			
	A. Defective communication (reasons not given)	2	3	2
	B. Expressive deficiencies (Especially: deficiencies of terminology and wordiness)	37	45	65
	C. Style deficiency	3	9	2
	D. Vague, unclear or ambiguous	14	15	21
	E. Too technical	5	4	5
	F. Too complex	2	4	5
	G. Hedge	2	3	2
	H. Too long	13	15	7
	I. Too brief	1	1	1
	J. Poorly organized	< 1	8	< 1
V.	PROBLEMS OF ROLE CONDUCT (Encroachment)	12	0	4
VI.	PROBLEMS OF SCIENCE AND PROFESSION (Especially: psychological theory, validity of tests, unsettled role of psychologist)	2	10	0

*Figures represent per cent of respondents in each category who made a negative comment or comments about psychological reports.

patient *is*" was jointly placed in the subcategories "exhibitionistic" and "not patient oriented." Such a procedure of necessity lends an impressionistic flavor to the data thus processed. To estimate the reliability of this method, therefore, the investigators, after a lapse of more than two months from the original sorting, independently sorted 200 randomly selected responses into the categories and subcategories already established. The level of agreement between the two investigators was 83%.

RESULTS AND DISCUSSION

The tabled data must be examined with two additional limitations in mind. Only about half of the respondents who returned research forms took the opportunity to make comments about psychological reports. There is no way of knowing whether those who failed to respond to this item had no negative feelings about re-

ports or did not complete the item for some other reason. In any event, the percentages shown in the table stem only from those who had any negative comment or comments to make, not of the entire professions as they are labeled in the table. Quite a few responses were positive in nature, some even laudatory of psychological reports. But this study deliberately attempted to seek out the negative features of clinical reports with a view to publicize them so that they might be reduced in future writing.

The principal issue which psychiatrists take with the psychological report is that it contains too much in the way of various forms of irresponsible interpretation. The data indicate many ways in which psychological material may be interpreted irresponsibly, but the tabled sub-category "assorted irresponsible interpretations" refers principally to such matters as overinterpretation, underinterpretation, insufficient objectivity, reflecting the examiner's personality, being textbookish, being not adequately related to theory, and just plain invalid interpretation. Psychiatrists are also greatly concerned with what is here called "expressive deficiencies", a term used principally to cover the use of objectionable terminology (primarily jargon) and wordiness. Many of them would be happier if more raw data and less speculation were included in the report. Many psychiatrists complain that the psychological report is unclear (or ambiguous) and too long. A sufficiently large number of psychiatrists to be reckoned with also seem to be concerned that the psychological report reflects encroachment upon other disciplines, primarily their own. The extent to which this is so might be greater than the figure in the table suggests since several items included under "Problem of Content", viz "offer diagnosis", and "offer recommendations" might well reflect concern with encroachment.

Perhaps the psychologist's greatest criticism of psychological reports centers about the expressive deficiencies, but he is about equally concerned that the psychological report might not be written with a practical or useful purpose in mind. The psychologists are critical that reports tend to be too general, that is, that they do not differentiate among patients, and they also tend to feel that various forms of irresponsible interpretation enter the report. Improper emphasis (overemphasis), particularly on psychoanalysis, pathology, diagnosis, dynamics, and improper emphasis (underemphasis), particularly on strengths and treatment possibilities, is yet another criticism. Like the psychiatrists, the psychologists are critical of unclear writing and excessive length. Unlike the psychiatrists, however, who complain of too little raw data in the report, the psychologists see the report as containing too much raw data. A number of psychologists seem to be cognizant of problems which center about the science and profession of psychology as opposed to problems of the individual clinician. In particular, such matters as deficiencies in supporting psychological theory, the validity of tests, and the unsettled role of the psychologist are mentioned.

The social workers join with both the psychiatrists and psychologists in being critical of expressive deficiencies. Two-thirds of the social workers who responded to the item "The trouble with psychological reports is . . ." made this complaint. Like their colleagues on the psychiatric team the social workers are also critical of irresponsible interpretation in the report and unclearness of writing. A number of social workers, like the psychologists and some psychiatrists too, feel that the report is not written with a practical purpose in mind.

Summary

A large number of psychiatric team members, psychiatrists, psychologists, and psychiatric social workers were asked to complete the sentence (or paragraph) root "The trouble with psychological reports is . . . " The comments were categorized into Problems of Content, Problems of Interpretation, Problems of Psychologist's Attitude or Orientation, Problems of Communication, Problems of Role Conduct, and Problems of Science and Profession. Some apparently meaningful patterns of criticism were found.

SOME CAUTIONS CONCERNING THE USE OF CHANGE SCORES

FRED HEILIZER

Veterans Administration Hospital, Northampton, Mass.

The use of initial scores as a control variable or selection device is a proper and useful statistical technique. The use of change scores to indicate improvement or deterioration of performance is also a proper and useful statistical technique. The simultaneous use of these techniques, however, is neither proper nor useful.

Churchill[1] presents a derivation for the correlation between initial scores and (a) difference scores, and (b) ratio change scores. The correlation between initial scores and difference scores is given by

$$r_{xd} = \frac{r_{xy} - \phi}{\sqrt{1 - 2r_{xy}\phi + \phi^2}} \qquad [1a]$$

where X = initial score, D = Y − X, and $\phi = \sigma_x/\sigma_y$. It can be seen that r_{xd} is a function of r_{xy} and ϕ. The correlation between initial scores and ratio scores is given by

$$r_{x\theta} = \frac{A\sqrt{1 + 3V^2}}{\sqrt{\frac{1 + 8V^2}{1 + 3V^2} + \frac{1 - r_{xy}^2}{\left(\frac{V}{W} - r_{xy}\right)^2}}} \qquad [1b]$$

where X = initial score, $\theta = (Y - X)/X$, $V = \sigma_x/M_x$, $W = \sigma_y/M_y$, and $A = -1$ if the quantity $(V - r_{xy}W)$ is positive and $+1$ if the quantity is negative. It can be seen that $r_{x\theta}$ is a function of r_{xy}, V, and W.

It is our purpose to provide a less complex derivation for the above formulae and to extend the derivation of [1a] to additional experimental situations by utilizing Standard Scores. (The additional derivations are not feasible for ratio change scores when Standard Scores are utilized, since ratios of infinity occur. This, in itself, is a pertinent comment on the use of ratio change scores.) In so doing, we make the assumption that a correlation is utilized solely to test an hypothesis of co-relation; it is not intentionally utilized to test an hypothesis of group differences on such parameters as means or variances. Therefore, the uses and limitations of a particular correlation are most properly indicated by utilizing variables with equated means and variances, *e.g.*, Standard Scores. In the presentation to follow, r refers to the general correlation formula while r' refers to the correlation formula which is derived by the use of Standard Scores.

If we are interested in the relationship between initial scores and difference scores, where X = initial score and D = Y − X,

$$r_{xd} = \frac{\Sigma XD/N - (M_x)(M_d)}{\sigma_x\sigma_d} \qquad [2a]$$

When X and Y are Standard Scores, $M_x = M_y = M_d = 0$ and $\sigma_x = \sigma_y = 1$. Therefore

$$r'_{xd} = \frac{\Sigma XD/N}{\sigma_d} = \frac{\Sigma[X(Y - X)]/N}{\sigma_d} = \frac{(\Sigma XY - \Sigma X^2)/N}{\sigma_d}$$

$$r'_{xd} = \frac{r_{xy} - 1}{\sigma_d}. \qquad [2b]$$

Formula [2b] indicates that when a correlational hypothesis is tested between initial scores and difference scores the correlation cannot be positive; it must be either zero or negative.

If we are interested in the relationship between initial scores and ratio change scores, where X = initial score and $\theta = (Y - X')/X$

$$r_{x\theta} = \frac{\frac{\Sigma X\theta}{N} - (M_x)(M_\theta)}{\sigma_x \sigma_\theta}. \qquad [3a]$$

When X and Y are Standard Scores,

$$r'_{x\theta} = \frac{\Sigma X\theta/N}{\sigma_\theta} = \frac{\Sigma\left[\frac{X(Y - X)}{X}\right]\Big/N}{\sigma_\theta} = \frac{\Sigma(Y - X)/N}{\sigma_\theta}$$

$$r'_{x\theta} = \frac{0}{\sigma_\theta}. \qquad [3b]$$

Formula [3b] indicates that when a correlational hypothesis is tested between initial scores and ratio change scores the correlation is zero.

If we are interested in the relationship between two sets of difference scores which contain a variable in common, where D = Y − X and E = Z − Y,

$$r_{de} = \frac{\frac{\Sigma DE}{N} - (M_d)(M_e)}{\sigma_d \sigma_e}. \qquad [4a]$$

When X, Y, and Z are Standard Scores,

$$r'_{de} = \frac{\Sigma DE/N}{\sigma_d \sigma_e} = \frac{\Sigma(XY + YZ - XZ - Y^2)/N}{\sigma_d \sigma_e}$$

$$r'_{de} = \frac{(r_{xy} + r_{yz}) - (1 + r_{xz})}{\sigma_d \sigma_e}. \qquad [4b]$$

Formula [4b] indicates that r'_{de} cannot be positive; it can only be zero or negative. Furthermore, if $r_{xy} = r_{yz} = r_{xz} = r$,

since $\sigma_d = \sqrt{\sigma_x^2 + \sigma_y^2 - 2r_{xy}\sigma_x\sigma_y} = \sqrt{2(1 - r_{xy})}$

and $\sigma_e = \sqrt{\sigma_y^2 + \sigma_z^2 - 2r_{yz}\sigma_y\sigma_z} = \sqrt{2(1 - r_{yz})}$

then [4b] becomes

$$r'_{de} = \frac{r - 1}{2(1 - r)}. \qquad [4c]$$

Formula [4c] indicates that $r'_{de} = -.50$ for all values of r, except when $r = +1$, in which case $r'_{de} = 0$.

A more complex situation results if we are interested in the relationship between difference scores and a control variable other than the initial score variable, where C = control score, and D = Y − X,

$$r_{cd} = \frac{\frac{\Sigma CD}{N} - (M_c)\ (M_d)}{\sigma_c \sigma_d}. \qquad [5a]$$

When C, X, and Y are Standard Scores,

$$r'_{cd} = \frac{\Sigma CD/N}{\sigma_d} = \frac{\Sigma[C(Y - X)]/N}{\sigma_d} = \frac{\Sigma CY/N - \Sigma CX/N}{\sigma_d}$$

$$r'_{cd} = \frac{r_{cy} - r_{cx}}{\sigma_d}. \qquad [5b]$$

Formula [5b] indicates that, as r_{cx} approaches + 1, it becomes increasingly more likely that r'_{cd} will be negative; as r_{cx} approaches − 1, it becomes increasingly more likely that r'_{cd} will be positive. It can also be seen that [2b] differs from [5b] only in the substitution of r_{cx} in the numerator of [5b] for *1* in the numerator of [2b]. Therefore, the more highly correlated are the initial and control scores, the more alike are the two formulae. Further, one might expect to obtain strikingly different values between r'_{cd}, on an *a priori* level, for widely differing values of r_{cx}.

The above derivations indicate that spurious correlations may result when change scores are utilized. When the change scores are correlated with a redundant variable, as in [2a], [3a], and [4a], the spuriousness seems obvious. However, when a redundancy may not be apparent, as in [5a], then careful attention must be paid to the hypothesis being tested and to the implications involved.

References

1. Churchill, E. Appendix in: Lacy, J. I. The evaluation of autonomic responses: toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, 67, 123-164.

THE EFFECT OF TIED RANKS ON THE RANK ORDER CORRELATION

FRED HEILIZER

Veterans Administration Hospital, Northampton, Mass.

The purpose of this paper is to indicate a potential bias in the rank order correlation when ties are involved. The notation used here is that used by Siegel[1].

The following formulae are appropriate for computing the rank order correlation when there are no tied ranks (r_s refers to the Spearman rank order correlation and, where X and Y are ranked variables, $x = M_x - X$, $y = M_y - Y$, $d = X - Y$):

$$r_s = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma d^2}{2\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{2\Sigma x^2 - \Sigma d^2}{2\Sigma x^2} = 1 - \frac{\Sigma d^2}{2\Sigma x^2}. \qquad [1]$$

Let us assume, for convenience, that the correction for tied ranks, $T = (t^3 - t)/12$), where t = the number of observations tied at a given rank, is the same in X and Y. Then, where $\Sigma x_c^2 = \Sigma x^2 - \Sigma T$, $r_s = 1 - (\Sigma d^2/2\Sigma x_c^2)$. Since $(\Sigma d^2/2\Sigma x_c^2) > (\Sigma d^2/2\Sigma x^2)$, the correction results in a change of all correlations toward the negative side of the r_s distribution, thus indicating that the effect of tied ranks, when no correction is used, is to bias r_s toward the positive side. The following recommendations are appropriate when tied ranks occur: (a) the correction should be applied when a positive correlation is significant but need not be applied when it is not significant; (b) the correction should be applied when a negative correlation is not significant but need not be applied when it is significant; (c) where the significance level is determined from a table of Σd^2, the occurrence of tied ranks invalidates the use of the table within the limits set forth in (a) and (b).

References

1. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.

HYPOCHONDRIASIS AND SOMATIC SEIZURE AURAS

BERNARD S. AARONSON

New Jersey State Hospital at Trenton

Application of the principle of stimulus generalization to personality data suggests that one source for the generation of traits may be the evocation of a response by a class of stimuli after the response has been elicited by a member of that class. Such a response should become more probable when any stimulus belonging to the same group appears at any level of intensity. In the case of epileptics, it would be anticipated that those whose seizures are preceded by gastric auras, paresthesias, or auras involving pain (somatic auras), should show more hypochondriacal concern than those with visual, auditory, psychic, kinaesthetic, or motor auras (nonsomatic auras), or no auras at all. Individuals with somatic auras should obtain profiles on the Minnesota Multiphasic Personality Inventory in which the *Hs* scale has a higher rank than in those obtained by individuals with nonsomatic or no auras at all.

Method

All MMPI profiles of epileptics were extracted from the files of the psychology department at New Castle State Hospital and all profiles with more than 20 items left unanswered or with any of the validating scores equalling or exceeding a T score

of 70 were eliminated. Individuals whose seizures occurred during sleep were also eliminated. The remaining sample was divided into somatic, nonsomatic, and no aura groups which were matched for age and sex since it has been shown[1, 2] that *Hs* is affected by both of these variables. This left three groups of six males and four females with mean ages of 28.9, 29.1 and 30.0 years in the somatic, nonsomatic and no aura groups respectively. These values are not significantly different. The scales were ranked in descending order and the rank of the *Hs* scale determined in each profile. All differences were tested by means of an analysis of variance of ranks.[3]

MMPI ranks rather than scale scores were used as the measure because the additional symptom of the aura might affect the absolute scale score. Ranks represent the relative strengths of the several response tendencies represented by the MMPI scales. An alteration in rank reflects not only an alteration in the frequency of responses emitted which are classifiable under a given scale name, but also an alteration relative to the frequency of responses classifiable under other scale names. Unless the absolute scale scores are of practically equal height, the ranks are less likely to be materially affected by the addition of one or two points to a scale score.

Results and Discussion

The mean rank on *Hs* of the somatic group was 4.4, of the nonsomatic group, 6.5, and of the no aura group, 7.0. It should be noted that only a ranking from one to ten is possible in a conventional MMPI profile. An analysis of variance of ranks over the whole data yields a chi square of 6.45, significant beyond the .05 level with two degrees of freedom. Comparison of the somatic and no aura group yields a chi square of 8.10, which is significant beyond the .01 level with one degree of freedom. A trend seems present for *Hs* to have a higher rank in the somatic than in the nonsomatic group, which does not attain significance at the .05 level. No difference seems present between the nonsomatic and no aura groups.

Epileptics with somatic auras seem to have more hypochondriacal concern than those with no auras. A trend seems present for them to show more somatic concern than those with nonsomatic auras, while the nonsomatic and no aura groups seem to differ only in a random manner. All of the patients in this study suffer from a chronic illness for which they have had to be hospitalized and all suffer from major seizures which come on only while they are awake. The only variable with respect to which they differ is the occurrence of the aura, which is a relatively mild experience compared to that which it foreshadows. The fact that this minor stimulus is associated with significant variation in the rank of the *Hs* scale suggests the role of stimulus generalization in the development of personality traits. A persistent symptom may elicit responses which then become more probable for the entire class of stimuli of which this stimulus is an instance.

Summary

Epileptics with somatic seizure auras obtain significantly higher ranks on the MMPI *Hs* scale than do those with no auras. A trend is also noted for those with somatic seizure auras to have higher ranks than those with nonsomatic auras. The nonsomatic and no aura groups do not seem to differ significantly. The role of stimulus generalization in the generation of personality traits is suggested.

References

1. Aaronson, B. S. Age and sex influences on MMPI profile peak distributions in an abnormal population. *J. consult. Psychol.*, 1958, *22*, 203-206.
2. Calden, G. and Hokanson, J. E. The influence of age on MMPI responses. *J. clin. Psychol.*, 1959, *15*, 1954-195.
3. Wilcoxon, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Company, 1949.

THE SHIPLEY-HARTFORD SCALE AND THE DOPPELT SHORT FORM AS ESTIMATORS OF WAIS IQ IN A STATE HOSPITAL POPULATION

LLOYD K. SINES AND HELEN SIMMONS

Fergus Falls State Hospital, Minnesota

Problem

The senior author[4] has reported the use of the Shipley-Hartford scale in estimating Wechsler-Bellevue scores in two VA populations. Further, several investigators (Himelstein[2], Olin and Reznikoff[3]), have evaluated the predictive accuracy of Doppelt's[1] short form WAIS in psychiatric settings. Because the conversion of Shipley scores to Wechsler equivalents among incoming patients at this hospital yielded consistent over-estimates when the tables developed in the VA settings were used, and in order to compare the efficiency of the Shipley scale with that of a validated "short cut" to WAIS estimation, the two methods were evaluated on a randomly selected sample of patients admitted to the Fergus Falls State Hospital.

Procedures and Subjects

All patients between the ages of 15 and 65 admitted to the hospital were screened for reading ability[1] and general (physical, psychiatric) amenability to psychological testing.

Approximately 45% of all patients admitted during a consecutive 15 month period met the criteria established, and were given a routine battery of psychological tests, one of which was the Shipley-Hartford scale, within the first week of hospitalization. From that group of patients a random sample of 30 was selected serially, and each patient was given a full (eleven subtest) WAIS *within one week* of the time at which the group tests (including the Shipley) had been completed.

The patients selected ranged in age from 16 to 62 (mean 40.3, sigma 13.5), and included 19 psychotics (12 schizophrenics), 4 neurotics, and 7 of miscellaneous diagnoses.

Results

WAIS full scale IQ ranged from 67 to 130 (mean 96.1, sigma 14.5), while Shipley total scores ranged from 13 to 70 (mean 37.0, sigma 17.6). Intercorrelations between the WAIS and Shipley subscales indicated that Shipley total score (uncorrected for omitted vocabulary items) was the best predictor of WAIS full scale IQ. The product moment correlation coefficient obtained between the latter two variables was .90, and the standard error of estimate in predicting WAIS IQ from Shipley total score was 6.33. Employing Wechsler's scheme of intellectual classification, it was found that the estimate of WAIS IQ made from Shipley total score correctly classified eighteen (60%) of the thirty subjects.

Doppelt short form IQ estimates were obtained, and the correlation between such estimates and actual WAIS full scale IQ was found to be .960. The standard error of estimate of IQ from the Doppelt formula was 4.06, while the accuracy of classification of the latter method was 67%. These figures are in substantial agreement with those reported by others in previous evaluations of the Doppelt formula.

Though estimation of WAIS full scale IQ by the Doppelt short form method tends to yield somewhat better predictions than does estimation from the Shipley-Hartford, the fact that scores on the latter scale are highly related to WAIS scores, and in addition, the fact that the scale requires no professional time for administration and scoring, justifies its use as a brief measure of intellectual functioning. Therefore, a conversion table for relating Shipley total scores (the best predictor) to

[1]The Stanford-Binet item *Reading and Report* (Form L, X, 2) was routinely administered. If the subject met the reading criterion for that item (not more than two errors), he was judged suitable for psychological testing requiring reading.

WAIS full scale IQ was constructed by solving the derived regression equation[2]. Table 1 provides the obtained values.

TABLE 1. WAIS FULL SCALE IQ EQUIVALENTS FOR SHIPLEY-HARTFORD TOTAL SCORES DERIVED FROM THE STATE HOSPITAL SAMPLE

S-H total score	WAIS full scale IQ	S-H total score	WAIS full scale IQ	S-H total score	WAIS full scale IQ
80	128	56	110	32	92
79	127	55	110	31	92
78	127	54	109	30	91
77	126	53	108	29	90
76	125	52	107	28	89
75	124	51	107	27	89
74	124	50	106	26	88
73	123	49	105	25	87
72	122	48	104	24	86
71	121	47	104	23	86
70	121	46	103	22	85
69	120	45	102	21	84
68	119	44	101	20	83
67	118	43	101	19	83
66	118	42	100	18	82
65	117	41	99	17	81
64	116	40	98	16	80
63	115	39	98	15	80
62	115	38	97	14	79
61	114	37	96	13	78
60	113	36	95	12	77
59	112	35	95	11	77
58	112	34	94	10	76
57	111	33	93		

SUMMARY

Patients between the ages of 15 and 65 who were admitted to the Fergus Falls State Hospital during a 15 month period were screened for reading ability and amenability to psychometric assessment. Thirty such patients who had taken a routine battery of psychological tests (including the Shipley-Hartford scale) were randomly selected, and were given a complete WAIS within one week following screening.

WAIS and Shipley scales were intercorrelated, and it was found that Shipley total score was the best predictor of WAIS full scale IQ ($r = .90$). Doppelt short form WAIS IQ estimates were also obtained and were found to correlate .96 with actual WAIS full scale IQ. Though it was concluded that the Doppelt formula provides more accurate estimates of WAIS IQ, in view of the very high degrees of relationship between Shipley and WAIS scores, and because of the ease and economy of administering and scoring the latter, a conversion table was provided whereby WAIS full scale IQ equivalents may be obtained for Shipley total scores.

REFERENCES

1. DOPPELT, J. E. Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. *J. consult. Psychol.*, 1956, *20*, 63-66.
2. HIMELSTEIN, P. Evaluation of an abbreviated WAIS in a psychiatric population. *J. clin. Psychol.*, 1957, *13*, 68-69.
3. OLIN, T. D., REZNIKOFF, M. The use of Doppelt's short form of the Wechsler Adult Intelligence Scale with psychiatric patients. *J. consult. Psychol.*, 1957, *21*, 27-28.
4. SINES, L. K. Intelligence test correlates of Shipley-Hartford performance. *J. clin. Psychol.*, 1958, *14*, 399-404.

[2] WAIS IQ = (.7463) ($S\text{-}H_{\text{total score}}$) + 68.455.

EDITORIAL OPINION

ACADEMIC PSYCHOLOGY: A SEMANTIC MORASS?

This editorial was stimulated by a chance remark of a prominent university psychologist who stated that he had counseled his son from taking psychology courses at the university while an undergraduate because he did not wish the son's high intuitive interest and understanding of behavior to be stultified by the particular academic orientation of that department. While it may be invalid to refer to the "academic orientation" of a department composed of many members presumably holding different viewpoints, there is nevertheless some truth in the observation that many departments have acquired reputations for the types of psychology taught there. One of the difficulties lies in the fact that all of the different systems of psychology have developed distinctive vocabularies, the comprehension of which constitutes a formidable exercise in semantics.

The student's plight is not relieved by the attempts of textbook writers to acquaint the beginner with all psychological systems in one bulging introductory text. Recently we have had occasion to review the content of the introductory psychology course at a leading college. In the first semester, the beginner was required to master the vocabularies of experimentalism, Behaviorism, Gestalt psychology, psychoanalysis, elementary statistics, and of a large number of illustrative experiments. By actual comparison, the 1959 introductory course was requiring a student to master materials given in the first graduate year 20 years ago and some of the examination questions were exact duplicates of Ph.D. preliminary examination questions of 20 years ago! Very few undergraduates are capable of comprehending much less digesting such an excercise in semantics except for those with very high memory ability who can parrot back the words.

At the rate at which complications of vocabulary are being introduced into academic psychology, future generations of psychologists will be required to devote their entire life spans to mastering the semantic intricacies of all the theoretical constructs in the field. Indeed the student is confronted with a semantic morass when he attempts to digest many systems as Hullian learning theory, Allport's event structure theory, Kelly's personal construct theory, Cattell's factorial theory, or modern decision theory, to cite a few of the worst offenders. We have here the dangerous situation where a student must devote more time to mastering the semantic complexities of a scientific vocabulary than he has available for directly experiencing the raw data of the science.

Human behaviors occur in almost infinite variety and it is not feasible to coin a new vocabulary to describe each new personality concept or clusterings of etiologic factors as may be discovered. To do so would intolerably deepen the semantic morass we are already entangled in. What is needed is a rigorous application of basic English to scientific communications of all types with the adoption of a basic vocabulary to which new terms would not be added except when no simpler suitable term was available. Journal editors could facilitate such a solution by rigorously cutting away semantic confabulations and excluding needless complexities. Indeed, this editor must confess that after 30 years in the field he is constantly encountering manuscripts which are comprehensible only by taking out an inordinate amount of time to laboriously study out the intricacies of some new semantic jungle which an author has staked out for himself. It might be very revealing to request a representative group of older academic psychologists to take comprehensive examinations covering *all* of the complexities of theory which have recently appeared.

The significance of all this for clinical psychology is that promising students in the clinical field should not be excluded from advanced study simply because they cannot master the semantic jungle which is academic psychology. There are many

students who have a genuine intuitive understanding of behavior, and who are by temperament suited for clinical work, but who may not have the semantic abilities to comprehend modern academic psychology. We would not advocate for a minute that academic psychologists be deprived of the right to pursue their semantic complexities. But we would insist promising clinical students should not be deprived of a career simply because they cannot master the semantic intricacies of academic psychology. It is entirely possible that schools of clinical psychology should develop their own curricula consisting of materials of proven clinical significance. Such curricula would expose the student to the complexities of academic theory but would not make promotion dependent upon ability to master them.

The situation would not be so serious if any eternal verities of genuine significance were being communicated in terms of such semantic obscurities since it could then be claimed that the student was being forced to "dig it out for himself". Too often, such is not the case. This editor has long since developed the habit of evaluating every sentence of a manuscript to discover whether anything of significance was being communicated and, if so, whether it is being expressed in the simplest operational terms and with a minimum of excess verbiage. If most of the current literature could be evaluated in such terms, it must be concluded that much semantic ado is being made over a lot of facts of very limited significance.

It should be the obligation of our senior colleagues to "boil down" the plethora of current semantic complications into the simplest operational terms. Almost 30 years ago, the writer struggled through the basic literature of Gestalt psychology, encountering a very difficult presentation semantically, but eventually coming to comprehend the basic contributions by translating them into simple operational terms. Indeed, the essence of Gestalt psychology is not the semantic complexities in which the original presentations were couched, but rather a series of ingenious experiments and observations which can be very simply stated. Similarly, the essence of more recent contributions is not in the semantic complexities in which authors seem to feel compelled to complicate the issues, but rather in a set of relatively simple operational viewpoints which can be set forth in simple operational terms.

The essence of psychology is behavior. The theories, concepts, operational methods and vocabularies which have developed in such complexities are strictly superfluous unless they deal with and communicate the issues both validly and economically. There are many factors operating on the current psychological scene tending to perpetuate and even to proliferate contemporary theoretical and semantic excesses. If our tremendously complicated field is to survive expeditiously, it must be streamlined by a rigorous paring down, at least on teaching and applied levels, of all excess baggage. Promising clinical students must not be failed because they cannot escape from semantic morasses.

F. C. T.

BOOK NOTICES

KOCH, SIGMUND (Ed.) *Psychology: A Study of Science.* Vol. I. New York: McGraw-Hill, 1959. Pp. 710. $9.75

This is number one of a seven volume series summarizing the major formulations of theoretical psychology, covering sensory, perceptual and physiological fields.

BASS, B. M. and BERG, I. A. *Objective Approaches to Personality Assessment.* Princeton, N. J.: Van Nostrand, 1959. Pp. 233. $4.95

STANDAHL, S. W. and CORSINI, R. J. (Eds.) *Critical Incidents in Psychotherapy.* Englewood Cliffs, N. J.: Prentice-Hall, 1959. Pp. 396. $6.95

Twenty-three authorities in the field of counseling and psychotherapy contribute critical incidents from their clinical practices which are evaluated and discussed by commenting consultants. Materials of this type should become important teaching aids.

TOMKINS, S. S. and MINER, J. B. *PAT Interpretation.* New York: Springer Publishing Co., 1959. $5.50. A manual for the Tomkins-Horn Picture Arrangement Test.

MIRA Y LOPEZ, E. *Myokinetic Psychodiagnosis.* New York: Logos Press, 1958. Pp. 186. $6.75

SZONDI, L., MOSER, U. and WEBB, M. W. *The Szondi Test.* Philadelphia: Lippincott, 1959. Pp. 309. $12.00

PATTERSON, C. H. *Counseling and Psychotherapy: Theory and Practice.* New York: Harper, 1959. Pp. 322. $5.00 The theory and philosophy of counseling from the viewpoints of phenomenonological psychology and client-centered therapy.

HARPER, R. A. *Psychoanalysis and Psychotherapy: 36 Systems.* Englewood Cliffs, N. J.: Prentice-Hall, 1959. Pp. 182, paper back. A remarkably terse and effective compendium of theories and systems.

ROTHNEY, J. W. M., DANIELSON, P. J. and HEIMAN, R. A. *Measurement for Guidance.* New York: Harper, 1959. Pp. 378. $5.00

LEVITT, M. (Ed.) *Readings in Psychoanalytic Psychology.* New York: Appleton-Century-Crofts, 1959. Pp. 413.

SARASON, S. B. *Psychological Problems in Mental Deficiency.* Third edition. New York: Harper, 1959. Pp. 678. $6.50

. . . A completely new volume in a distinguished series

THE FIFTH

MENTAL MEASUREMENTS YEARBOOK

Edited by Oscar Krisen Buros
Director, Institute of Mental Measurements
Rutgers, The State University

- Comprehensive, up-to-date information on 959 tests and assessment techniques of all kinds—group and individual; educational, psychological, vocational.
- 698 critical test reviews by 350 contributing authorities.
- A 237-page section on projective and nonprojective character and personality tests.
- More than 6,000 references on the construction, validation, use, and limitations of specific tests.
- Hundreds of concise excerpts from journal reviews of books on measurements and related fields.

The Yearbook is an invaluable aid to anyone who must evaluate, select, use or interpret tests and test results. Completely indexed and cross-referred to previous volumes. Pp., xxx, 1292. $22.50 plus postage. We pay postage when payment accompanies order. Order your Fifth Yearbook directly from

THE GRYPHON PRESS 220 Montgomery Street, Highland Park, New Jersey

COMPANION VOLUMES

PRINCIPLES OF PSYCHOLOGICAL EXAMINING

by

FREDERICK C. THORNE, M.D., Ph.D.
Editor, Journal of Clinical Psychology

PRICE: $7.50

PRINCIPLES OF PERSONALITY COUNSELING

by

FREDERICK C. THORNE, M.D., Ph.D.
Editor, Journal of Clinical Psychology

PRICE: $6.00

MENTAL DEFICIENCY

IN RELATION TO PROBLEMS OF GENESIS, SOCIAL AND OCCUPATIONAL CONSEQUENCES, UTILIZATION, CONTROL AND PREVENTION
by J. E. WALLACE WALLIN, Ph.D.

This important book summarizes the author's life-long experience in dealing with practical social problems relating to the control of mental defectives. Dr. Wallin bases his authoritative opinions upon an intensive review of the literature in this field and his conclusions reflect the latest scientific information. This book differs from other works on mental deficiency in that it deals with many practical issues for which society urgently demands some answers. It should be read by all psychiatrists, psychologists, social workers, educators, parents and other workers in the field.

PRICE $5.00 - - - ORDER NOW

THE HOWARD INK BLOT TEST

A NEW PROJECTIVE TECHNIQUE

- Derived experimentally by James W. Howard, Ph.D.
- An individual test consisting of 12 plates 8½" x 11". The blots are somewhat larger than the Rorschach and with a wider range of stimulating features. Blots of varied colors selected from an experimental sample of 10,000 blots.
- Not intended to be a parallel test to the Rorschach. Statistical findings from a "normal" group show the test to differ markedly from the Rorschach in ways which should indicate greater sensitivity.
- Selection and arrangement of blots made on the basis of evidence of character, degree and spread of stimulation.
- Data from normative sample indicate more frequent light determined responses, higher incidence of color responses with different C, CF and FC relationships, more movement responses and fewer animal responses. Greater diagnostic sensitivity with wider range of response.

PRICE: $12.50 (In U. S. and Canada; foreign postage $.50 extra)

HOWARD INK BLOT TEST MANUAL

MONOGRAPH SUPPLEMENT NO. 10 FROM JULY 1953 ISSUE

Price $2.00

Order From

JOURNAL OF CLINICAL PSYCHOLOGY
5 Pearl Street Brandon, Vermont

Clinical Psychology Monographs

NO. 1 F. Y. BILLINGSLEA
Bender Gestalt Test Manual — $ .90

NO. 2 CARTER AND BOWLES
Qualitative Aspects of Psychological Examining — $1.25

NO. 3 SYMPOSIUM ON
Training in Clinical Psychology — AVAILABLE ONLY IN COMPLETE SETS

NO. 4 SYMPOSIUM ON
Critique of Nondirective Therapy — $1.25

NO. 5 JOHN T. BUCK
The House-Tree-Person Test Manual — $2.00

NO. 6 BEN KARPMAN'S
Objective Psychotherapy — $2.00

NO. 7 SYMPOSIUM - - JAN. 1950
Statistics for the Clinician — $2.00

NO. 8 ROBERT TYSON - - JAN. 1951
Current Mental Hygiene Practice — $2.00

NO. 9 SYMPOSIUM OF RECENT CONTRIBUTIONS TO THE FIELD - - APRIL 1953
Counseling the Mentally Defective — $1.25

NO. 10 JAMES W. HOWARD - - JULY 1953
The Howard Ink Blot Test — $2.00

NO. 11 ALBERT ELLIS - - July 1955
New Approaches to Psychotherapy Techniques — $2.00

NO. 12 GARDNER LINDZEY et al - - April 1959
The Thematic Apperception Test: — $2.50
An Interpretive Lexicon

NO. 13. DELL LEBO - - May 1959
Guide Book for Thesis Composition — $1.25

Order Now

JOURNAL OF CLINICAL PSYCHOLOGY

5 Pearl Street, Brandon, Vermont

2 NEW TITLES IN AMERICAN LECTURES IN PSYCHOLOGY

Edited by

MOLLY HARROWER, Ph.D.

Research and Consulting Psychologist, New York
Professor of Research in Clinical Psychology in the Department of Psychiatry
Temple University School of Medicine
Philadelphia, Pennsylvania

A FUNCTIONAL APPROACH TO TRAINING IN CLINICAL PSYCHOLOGY: Via Study of a Mental Hospital *by* **Abraham S. Luchins,** *Univ. of Miami.* This book is a positive NO to the practice of turning the psychologist trainee loose in the hospital with instructions to "observe." Its broad orientation operates against freezing the psychologist in his traditional role of technician. A stimulating and different view of a training program for the clinical psychologist in a mental hospital. EMPHASIZES THE ROLE OF THE PSYCHOLOGIST AS A STUDENT OF BEHAVIOR AND SOCIAL PROGRESS. Doctor Luchins puts into focus ways to get acquainted with a mental hospital, not only inside but even outside the hospital walls. The trainee sees the whole hospital intimately rather than a part well before he ever takes the Rorschach cards out of the kit. Quoting a reader . . . "Luchins' program gives students a healthier respect for the kind of problems which nurses, aides, social workers, and psychologists have to face in a hospital. And once having realized this, and gotten to know these other hospital people, it would seem that test reports would contain less private jargon and more useful, shareable observations."

Pub. Sept. '59, 334 pp., $7.50

1

DRIVE STRUCTURE AND CRIMINALITY: Criminobiologic Investigations *by* **Hans Walder,** *Attorney General, Zurich. Edited* by **Lipot Szondi** *and translated* by **Marvin W. Webb.** By means of the most highly developed techniques of criminal investigation and personality exploration employed in Western Europe, the psychodynamics of criminality are pursued to their innermost recesses within the human psyche. The global personality of the offender is gleaned upon a longitudinal basis, chiefly through the application of the Szondi test, a technique favored by Continental jurists. The Szondi test is particularly instructive in portraying the unique constellations of drive traits and tendencies which are interrelated to the several types of offenses. It is clearly demonstrated that a given offender is capable of committing only certain categories of offenses and not others. This work fosters improved methods of crime detection. It seeks to illumine the more difficult but more significant areas of punishment, reformation, and crime prevention. Doctor Hans Walder meets in full measure the criterion that this comprehensive theory of criminogenics shall be presented by a criminologist of repute.

Pub. Aug. '59, 192 pp., 164 ill., $7.50

2

Send for our complete list of titles in American Lectures in Psychology

CHARLES C THOMAS • PUBLISHER
301-327 East Lawrence Avenue
SPRINGFIELD • ILLINOIS

RESEARCH AND TRAINING

THE PROGRESS MADE by the children entrusted to Devereux's care is the first concern of every professional member of the Schools' staff.

Devereux's program of evaluation, therapy, and education gives excellent opportunity for research and for the training of professional personnel.

The Institute for Research and Training of the Devereux Foundation provides pre-doctoral and post-doctoral fellowships, internships, and traineeships in psychology, psychiatry, psychiatric social work, and education. These are designed to contribute to the understanding of the care, education, training, and treatment of the emotionally disturbed and the mentally retarded child in a residential setting.

Professional inquiries about the program should be addressed to Dr. Henry Platt or Dr. George Spivack, Director of Research, Devereux Schools, Devon, Pennsylvania.

CLINICAL STAFF

J. Clifford Scott, M.D.
Edwin H. Abrahamsen, M.D.
Aurelio Buonanno, M.D.
Charles M. Campbell, Jr., M.D.
Fred J. Culeman, M.D.
Ruth E. Duffy, M.D.
William F. Haines, M.D.
Herbert H. Herskovitz, M.D.
Robert L. Hunt, M.D.
Richard H. Lambert, M.D.
Leonardo Magran, M.D.
Joseph J. Peters, M.D.
Alvis J. Scull, M.D.
Jacob S. Sherson, M.D.
Albert S. Terzian, M.D.
Walter M. Uhler, M.D.
Tirso L. Vinueza, M.D.

Lance Wright, M.D.
F. Ellsworth Henry, S.T.D.
Milton Brutten, Ph.D.
William J. Cohen, Ph.D.
Dorothy E. Conrad, Ph.D.
Sidney L. Copel, Ed.D.
Michael B. Dunn, Ph.D.
Shirley M. Jahnson, Ph.D.
John R. Kleiser, Ph.D.
Murray Levine, Ph.D.
Henry Platt, Ph.D.
Edgar A. Smith, Ed.D.
George Spivack, Ph.D.
Herbert A. Sprigle, Ph.D.
Anne Howe, M.S.
Kenneth E. Evans, B.S.
G. Henry Katz, M.D.
Psychoanalytic Consultant

THE DEVEREUX FOUNDATION

A nonprofit organization — *Santa Barbara, California*

Founded 1912 — *Devon, Pennsylvania*

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX
Administrative Consultant
EDWARD L. FRENCH, Ph.D.
Director
WILLIAM B LOEB
Treasurer

Professional inquiries should be addressed to JOHN M. BARCLAY, *Director of Development*, or CHARLES J. FOWLER, *Registrar*, Devereux Schools, Devon, Pennsylvania; western residents address KEITH A. SEATON, *Registrar*, Devereux Schools in California, Santa Barbara, California.